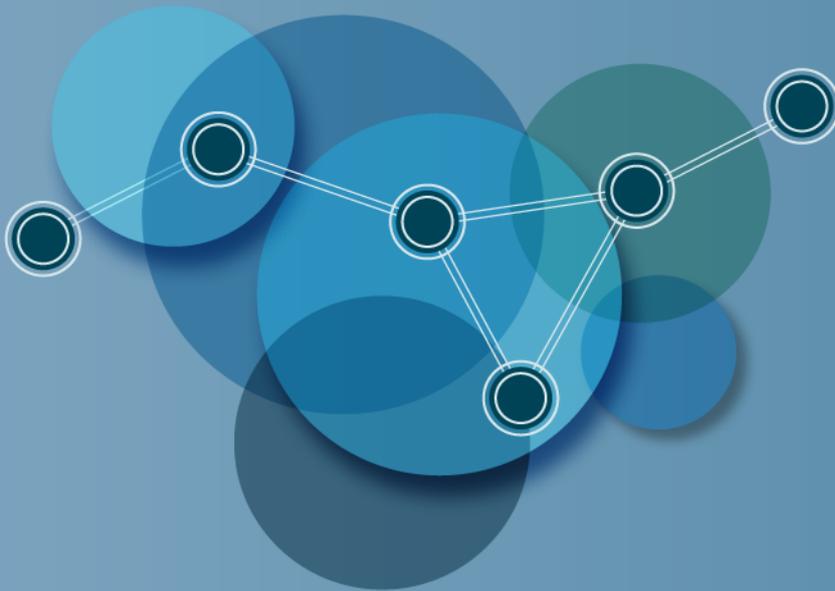François Habryn

CUSTOMER INTIMACY ANALYTICS

Leveraging Operational Data to Assess Customer
Knowledge and Relationships and to Measure their
Business Impact

KIT Scientific Publishing

François Habryn

Customer Intimacy Analytics

Leveraging Operational Data to Assess Customer Knowledge
and Relationships and to Measure their Business Impact

Customer Intimacy Analytics

Leveraging Operational Data to Assess Customer
Knowledge and Relationships and to Measure
their Business Impact

by
François Habryn

Customer Intimacy Analytics

Leveraging Operational Data to Assess Customer Knowledge and Relationships and to Measure their Business Impact

Zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der Fakultät für Wirtschaftswissenschaften
des Karlsruher Institut für Technologie (KIT)

genehmigte

Dissertation

von

François Habryn

Tag der mündlichen Prüfung: 16. Februar 2012
Referent: Prof. Dr. Gerhard Satzger
Korreferent: Prof. Dr. Rudi Studer

Karlsruhe

Preface

In today's economies, firms are characterized by an increasing degree of service orientation. Long-term customer relationships and individualized solutions are emphasized– up to the point where the willingness and ability to focus on "customer intimacy" turns into a unique type of strategy, as an alternative to product leadership and operational excellence (Treacy & Wiersema, 1993). Unfortunately, currently available CRM systems hardly support customer intimacy based strategies as they mostly focus on discrete sales transactions. However, the emerging area of business analytics offers IT-based concepts, methods, and tools that may open up a huge potential for firms to exploit existing customer interaction data as well as to augment and improve their CRM approach. They actually can provide tremendous analytical support both for designing and implementing customer intimacy strategies and for monitoring their effectiveness.

The work of François Habryn takes on this opportunity and makes significant and innovative contributions along three dimensions. Firstly, it decomposes the notion of "customer intimacy" into operationally meaningful and measurable components - as a prerequisite for an analytical evaluation of the quality of customer relationships. Secondly, it develops metrics based on existing interaction data to be applied to these components. And thirdly it provides a methodology and even a fully-fledged tool to test the ability of these metrics

to actually reflect relevant customer intimacy in practice. The results François Habryn obtained in a real case scenario are convincing as is the positive feedback that he has received at academic conferences as well as from various industry partners.

The work is a truly remarkable example for the capabilities of inter-disciplinary approaches to create innovative solutions to problems: François Habryn addresses the challenge of assessing customer intimacy from both the managerial and IT perspectives by integrating concepts grounded in relationship marketing, strategic management, business analytics, social network analysis, and software engineering. The insights gained should be highly relevant for leaders and managers in strategy, marketing, and sales in service-oriented companies as well as for consultants and IT providers in the CRM space.

I wish the audience an inspiring, enjoyable, and fruitful reading of this book and hope that this work will see the distribution in academia and industry that it deserves.

Prof. Dr. Gerhard Satzger
Director IBM Business Performance Services Europe

Acknowledgements

I would like to express my sincere gratitude to all those who helped me during the course of this thesis with their advice and support. First and foremost, I would like to thank Prof. Dr. Gerhard Satzger for consistently supporting me throughout all the phases of this project, for always being available to provide me with sound advice, as well as for leading my research to a high quality. I also wish to acknowledge the opportunity given to me by IBM in November 2007 to participate in the creation of the Karlsruhe Service Research Institute (KSRI) and to complete this doctoral thesis. This was a fantastic experience and I am grateful to Gerhard Satzger, Martin Jetter, and IBM for this opportunity.

I also wish to express my gratitude to Prof. Dr. Rudi Studer for being the second reviewer of this thesis. His valuable advice, support, and friendly encouragement assisted me greatly in completing this work. I am also very grateful to Prof. Dr. Hagen Lindstädt and Prof. Dr. Thomas Lützkendorf for accepting to be part of the examination board as Examiner and Chairman, respectively, and for their constructive advice and comments.

This work would not have been possible without the support of all my colleagues at KSRI. I would like to show gratitude in particular to my colleagues in the Service Innovation and Management team with whom I spent four excellent years: Prof. Dr. Hansjörg Fromm,

Andreas Neus, Robert Kern, Axel Kieninger, Peter Hottum, Marc Kohler, and Johannes Kunze von Bischhoffshausen. In addition, I wish to thank Dr. Benjamin Blau, Dr. Arun Anandasivam, Dr. Jeroen Schepers, and Gielis von der Heijden who advised me in the initial and final phases of my thesis. I also wish to acknowledge the KIT students who completed their diploma, bachelor, and master thesis under my supervision: Thomas Herzig, Lukas Lampe, and Hakan Bilgic, whose skills and dedication to the project were invaluable.

I would like to show appreciation to CAS Software AG (CAS) with the help of whom I was able to implement the prototype *CI Analytics* and to perform a survey which allowed the overall validation of this thesis. I would like to express my gratitude to Dr. Bernhard Kölmel for actively supporting this work within CAS, to Martin Hubschneider for allowing me to perform this project in his company, as well as to all the CAS employees who participated in the survey.

Finally, I would like to express thanks to those who provided me the most precious assistance. I owe my deepest gratitude to my parents for the environment in which I grew up, for their constant encouragement, and for always being there when I needed them. I also wish to give a very special thank you to Anna for her care and patience. Anna gave me confidence when I was in doubt and encouraged me when I was in low spirits.

François Habryn

Abstract

The ability to capture customer needs and to tailor provided solutions accordingly, also defined as customer intimacy, has become a significant success factor in the Business to Business (B2B) space – in particular for increasingly "servitizing" businesses. This growing importance of customer intimacy is driven by a fast development of the service industry, higher expectations on the demand side, and a shift in the role of the customer from passive value receiver to active value co-creator. However, the measurement and management of customer intimacy lacks analytical support. Even though customer relationship management (CRM) systems are well established today, they do not yet provide the appropriate means for supporting the implementation of a customer intimacy strategy. So far, customer intimacy was not given the adequate focus from the IT perspective and, thus, many organizations still struggle with measuring and proactively managing the degree of customer intimacy established with their customers.

In the scope of this thesis, the solution *CI Analytics* has been conceived, implemented, and validated in order to remedy this issue. *CI Analytics* complements existing CRM systems with the capability to assess and monitor the degree of customer intimacy established by a provider with its customers in a B2B context. It applies business analytics and social network analysis technology in order to provide

an accurate, real-time, and easily implementable assessment of customer intimacy with two levels of analysis: the individual level and the organizational level. *CI Analytics* leverages customer related data which is available in the information system of the provider (such as interactions, projects, and sales records) to derive customer intimacy metrics. These metrics are subsequently used to infer the established customer intimacy as well as its impact on business results.

Multiple benefits can be derived from the solution proposed by this thesis. First, *CI Analytics* allows an organization to benchmark the effectiveness of its customer intimacy strategy with different customers and, thus, supports this organization with regard to its customer investments. Second, this solution provides a systematic graph-based overview of the interactions among provider and customer employees, as well as a visualization of their evolution over time, thereby enabling the provider to proactively act upon any changes in the activity and interaction patterns with the customer. Finally, *CI Analytics* fosters the exchange of customer knowledge among the provider employees by facilitating the identification of employees inside the organization who acquired some specific customer knowledge and established relationships with customer employees.

The solution *CI Analytics* has been prototypically implemented in order to validate the feasibility of the proposed customer intimacy assessment and monitoring. This software allows different users in the provider organization to visualize in real time the investments performed by the provider employees in terms of interaction time in order to acquire customer knowledge and to establish relationships with customer employees. In addition, this software graphically represents the business impact of the customer intimacy strategy for specific customers and for specific time frames by means of dedicated customer intimacy performance indicators. *CI Analytics* has been evaluated in an enterprise setting with real data from the IT software and service provider CAS Software AG. This evaluation confirms the relevance of the proposed solution as well as allows the organization to gain insights on the patterns of interactions leading to a successful acquisition of customer knowledge and to an effective establishment of high-quality customer relationships.

Contents

List of Figures

List of Tables

Part I.

Foundations and Preliminaries

1. Introduction

A recent survey conducted in 2010 with 1500 chief executive officers worldwide established that, today, successful organizations "make customer intimacy their number-one priority" (IBM Institute for Business Value, 2010, p.9). Customer intimacy has gained momentum over the last years as it is perceived as a means to develop a sustainable business strategy in mature markets such as Europe and the United States, which are characterized by limited growth, fierce competition, and demanding customers.

Customer intimacy was first introduced by Treacy & Wiersema (1993) as one of three value disciplines, along with operational excellence and product leadership that, if executed well, leads to market leadership. Several firms in various industries, including IBM, Panalpina, Unilever, and General Electric Healthcare, were influenced by the concept of customer intimacy. Defined as the ability to "continuously tailor and shape products and services to fit an increasingly fine definition of the customer" (Treacy & Wiersema, 1993, p.87), customer intimacy determines an organization's business strategy and, as such, critically impacts its operations and performance (Hambrick, 1980).

In the contemporary challenging business environment, multiple companies strive to find new sources of growth and, thus, strengthen their relationships with customers in order to achieve new forms

of competitive advantages (Tuominen *et al.*, 2004; Day, 2003). In that regard, a firm which successfully pursues customer intimacy derives strategic benefits from its knowledge of, and relationship with, customers. For instance, the customer intimate firm proactively improves its value proposition and becomes its customers' preferred partner by customizing its offering to their specific requirements (Wallenburg, 2009). This firm embeds its customers in the value creation process and leverages their ideas in order to conceive innovative solutions. The firm adhering to customer intimacy also increases the loyalty of its customers, thereby protecting its investments by establishing long-term relationships.

Customer intimacy is particularly relevant in a service context as the development of a customer intimacy strategy covers two essential characteristics of services, namely the individualization of the offering to customer needs, and the intensification of the customer interactions in order to co-create value with customers (Bruhn & Georgi, 2006). Numerous companies that were known for their product centered portfolios have developed services-focused business models. For instance, Rolls-Royce and IBM, which used to generate over 60% of their revenues in 1995 with products, redesigned their offerings and realized in the past three years over 55% of their revenues with services. This transformation of the firm's business model from selling goods to selling solutions including goods and services is called servitization (Vandermerwe, 1988; Neely, 2009). Customer intimacy is potentially an adequate type of business strategy for companies undergoing a servitization endeavor and which try to strengthen their relationships with customers and to individualize their offerings.

Pursuing a customer intimacy strategy poses some specific challenges to the organization. In order to implement a customer intimacy strategy, the provider needs to manage the relationships established with customers as well as the acquired knowledge related to customers. In that regard, this thesis focuses on the specific challenges of business to business (B2B) markets. In a B2B context, both the provider and the customer consist of multiple teams and individuals. On the customer side, this means that, in most cases, users and purchasers are

different individuals inside the customer organization. While the decision to select one or the other B2B provider is made by purchasers, the users actually get in contact with the provided solution, and as such assess its quality and performance. Thus, the B2B provider must consider the needs and requirements of the different stakeholders inside the customer organization in order to successfully manage the relationship with the customer organization and successfully implement its customer intimacy strategy (Homburg & Jensen, 2004).

On the provider side, the ongoing servitization has a substantial impact on the organization, blurring the boundaries between sales, services, marketing, and even manufacturing departments (Oliva & Kallenberg, 2003). Sales employees, who were spokespersons for the firm's products have become sales consultants who understand and solve customer problems, leveraging knowledge and expertise across the entire provider organization (Sheth & Sharma, 2008). Reciprocally, service employees become increasingly involved in the selling process as they develop unique means to gather customer knowledge and understand the customer's mindset. Thus, managing customer relationships and pursuing customer intimacy in a B2B context requires the provider to thoroughly manage the complex and dynamic social network resulting from the interactions of his employees with customer ones. This development drives the need to redesign the interfaces among the internal departments of the B2B provider, "in terms of structure, communication patterns, information sharing, collaboration, and strategic outcome" (Biemans *et al.*, 2010, p.183). Zack *et al.* (2009, p.402) confirm that "firms achieving high customer intimacy engaged in the widest range of knowledge management practices."

From an academic perspective, customer intimacy overlays, in part, with prominent marketing concepts such as relationship marketing (Berry, 1983) and the modern perspective on services, namely the service-dominant logic (Vargo & Lusch, 2004a). Relationship marketing and the service-dominant logic take their root in a paradigm shift that positions the relationship with the customer as a central determinant of the marketing strategy, rather than the delivery of the product or service itself. Grönroos (1994), for instance, contrasts

the "4Ps marketing mix" of the transactional marketing which is dominated by the quality of the output and measured by market share, with relationship marketing, which is driven by the quality of the customer interactions and individually measured with each customer. Vargo & Lusch (2008b) qualify the service-dominant logic as focused on the exchange of knowledge and skills among partners rather than on the exchange of tangible goods, thereby contrasting the service-dominant logic with the goods-dominant logic. As it will be explained in chapter 2, customer intimacy is rooted in the concept of relationship marketing and shares several commonalities with the service-dominant logic.

1.1. Research Problem

From the IT perspective, the choice to pursue the value discipline customer intimacy directly impacts the IT governance of the organization and its infrastructure design. Weill & Ross (2004) investigated the influence of customer intimacy on IT governance by means of a survey with 250 enterprises worldwide. They concluded that customer intimacy driven organizations "strive for a single view of the customer", require analytical tools "to expose customers with the greatest lifetime value", and "implement customer relationship management (CRM) systems to support data standardization" (Weill & Ross, 2004, p.164). CRM systems aim at enabling to collect vast amounts of customer data and to constructively analyze, interpret, and utilize it (Payne, 2005). Such systems, therefore, support the development of a customer intimacy strategy. Several sources confirm that CRM systems have been widely adopted in order to achieve this objective. A recent Gartner report estimates the size of the CRM application market over $10 billion (Maoz *et al.*, 2010). Sackmann *et al.* (2008) found that, in 2008, 68% of the 292 German enterprises they surveyed had already implemented a CRM solution, and another 20% were planning to do so in 2009.

However, some evidence leads to question the actual benefits of CRM systems and in particular their positive association with business performance (Reinartz *et al.*, 2004). While Kale (2004) estimated the

CRM project failure rate between 60% and 80%, Dickie (2007) evaluated that only 20% of the organizations generated additional revenues from their CRM investments. Even though the customization of products and services is established as a value driver for the adoption of CRM, several CRM projects solely lead to an improvement of sales force efficiency and effectiveness (Richards & Jones, 2008). Blois (2008, p.1) states that "(CRM) software on the market today helps automate processes, but does not necessarily provide incremental value back to the user." He also considers that CRM systems are only used to track the progression of the sales opportunities from initial leads towards contracts (Blois, 2008).

In order to explain this phenomenon, Liang (2009) considers that IT systems have been so far adopted with transactional focus and operational excellence in mind. He argues that "the role of customer intimacy has been under-investigated" from the IT perspective (Liang, 2009, p.1). Even though CRM systems aim at managing customer related data, the customer knowledge which is derived from this data is mostly limited to the transactional perspective. The CRM system helps answering questions such as which products have been sold, in which quantities, when and by whom. However, more complex questions related to the needs of the customer, his future plans, or his purchasing behavior hardly find an answer in such systems. A survey performed with 122 senior executives in Western Europe acknowledges that firms' knowledge management capabilities are the weakest when knowledge is related to customers: "Despite the heavy investments firms have made in CRM systems in recent times, only 23% of the surveyed executives say they are effective in capturing and exploiting information on customer preferences and behavior" (Ernerst-Jones, 2005, p.7).

Considering the employees' perspective, this survey also indicates that organizations particularly struggle with exploiting knowledge of their employees (Ernerst-Jones, 2005). It is most likely that some provider employees who have spent time working for, and interacting with, the customer know the customer processes, how decisions are influenced and taken, and how budget is made available in the customer organization. These employees know how to effectively bring

new ideas inside the customer organization and, reciprocally, how to obtain useful feedback from the customer. They are also aware of the customer employees that favor their own organization and those who favor the competitors. In short, these provider employees know how to manage the three types of customer knowledge proposed by Gibbert *et al.* (2002): *about* the customer, *from* the customer, and *for* the customer. Thus, these employees have developed a certain degree of customer intimacy with the customer and the customer employees. However, because customer knowledge is often tacit and quickly outdated, provider employees do not have the means to store it in an explicit manner in the CRM system, and the provider does not have the capability to assess the degree of customer intimacy established by its employees with its different customers.

At the organizational level, the customer contribution margin is the most basic conception for assessing the profitability of business relationships (Wengler, 2006). However, an empirical analysis performed in 2006 reveals that only 30% of the surveyed organizations take this parameter into account, and 80% of them solely use transaction volumes in order to rate their customers (Wengler *et al.*, 2006). Taking the broader perspective of customer intimacy, a thorough literature review (Habryn *et al.*, 2010) which is further refined in section 4.1 of this thesis acknowledges that, as of today, there is no operational means for an organization to assess the degree of customer intimacy established with customers.

As a result, the central problem which is investigated in the scope of this thesis is concerned with the lack of easily exploitable solutions for an organization to assess the degree of customer intimacy that it has established at both the individual and organizational levels with its customers.

1.2. Research Objective

In order to address the issue presented in the previous section, the objective of this thesis is to develop a solution for assessing and monitoring the degree of customer intimacy established by an organization with its customers.

As illustrated in figure 1.1, the various interactions and activities of the provider employees with customer employees lead to the establishment of different degrees of customer intimacy between entities of the provider and customer. For instance, it is most likely that provider employees who worked on a customer project at the customer location developed a higher customer intimacy than other employees who only had limited interactions with the customer: they gathered more knowledge about the customers as they spent time with its employees and used this knowledge to adapt the solution they developed. The different business units, teams, and employees of the provider, thus, established different degrees of customer intimacy with the business units, teams, and employees of the customer. In order to analyze the degree of customer intimacy between the provider and the customer, it is therefore necessary to drill down the analysis to multiple levels of details.

Consequently, the assessment of the degree of customer intimacy should be performed in the scope of this thesis at two levels of granularity: the *organizational* level and the *individual* level. The organizational level indicates the customer intimacy established with customer organizations and its entities such as teams and business units. The individual level refers to the degree of customer intimacy established by provider employees with customer employees.

In order for this customer intimacy assessment to be relevant and usable by a provider, it needs to be up-to-date, accurate, and easily implementable. Making up-to-date assessments is a particularly challenging task as the information related to customer intimacy changes rapidly. For instance, customer needs may quickly evolve after a strategic reorientation. The customer might change its purchasing policy or decide to develop a new market for which he has new requirements. In addition, the customer organization and structure are also modified on a regular basis. If some customer employees with whom the provider had established qualitative relationships take a new position, the provider's ability to access knowledge about the customer and influence the customer may decrease, thereby impacting the customer intimacy established with the customer. The pro-
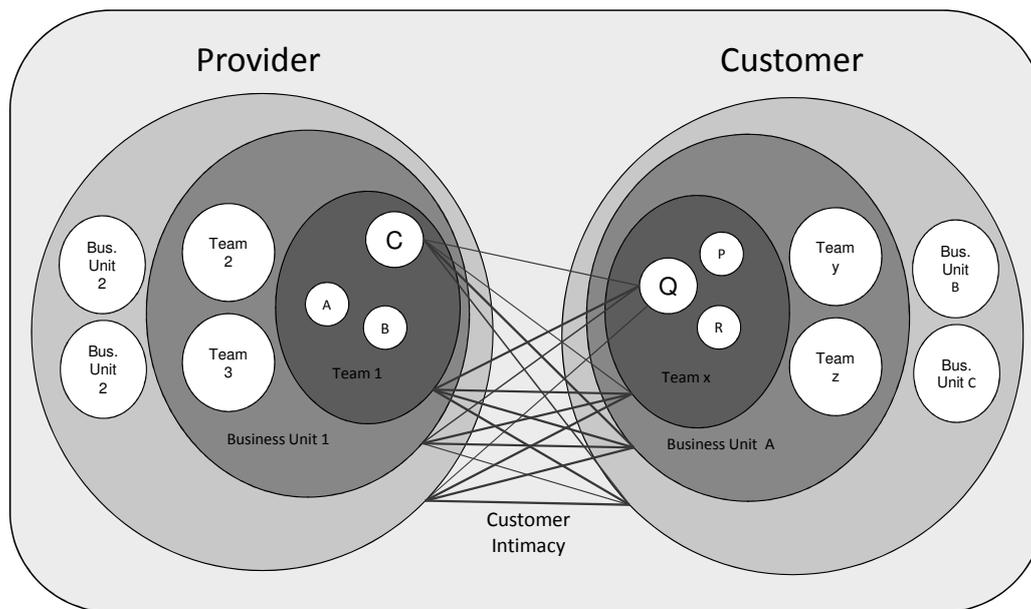
Figure 1.1.: Different Degrees of Customer Intimacy Between Provider and Customer Entities

vider must, therefore, have up-to-date information on such changes in order to successfully implement its customer intimacy strategy.

In order to obtain accurate information, this degree of customer intimacy should be evaluated across all departments of the provider organization. Indeed, the lack of information for specific provider employees or teams might lead to wrong interpretation of the degree of customer intimacy and restrain the ability to disseminate customer knowledge inside the provider organization. If an provider employee who has a very strong insight about the customer is not identified, his colleagues cannot benefit from his knowledge.

Finally, to achieve an easily implementable solution, the approach should not impact the provider employees with significant additional workload and should integrate seamlessly with the existing IT environment. If the provider employees have to spend a lot of time to enter customer intimacy related data into the system, they will be reluctant to using this solution. For these reasons, the customer intimacy assessment should be performed as far as possible automati-

cally, in a real-time fashion, and it should leverage readily available data.

1.3. Research Approach

In order to fulfill the requirements outlined in the previous section, this thesis is grounded in the areas of business intelligence and business analytics. The notion of business intelligence has been given multiple meanings in past literature. Turban *et al.* (2011, p.19) suggest a broad interpretation of business intelligence and define it as "an umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies [...] to enable interactive access to data, to enable manipulation of data, and to give [...] the ability to conduct appropriate analysis." According to Turban *et al.* (2011), business analytics is a part of business intelligence which is explicitly concerned with the exploitation of data by business users by means of either simple reports and queries or sophisticated mathematical and statistical methods such as data-mining. Davenport & Harris (2007, p.7) acknowledge that analytics is a subset of business intelligence and define it as "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions."

Business intelligence and in particular business analytics have received a growing interest over the past years. They are perceived as the means to take informed decisions upon the vast amount of gathered data and as a new source of competitive advantages (Davenport, 2006). An extensive study performed with 4500 managers and executives acknowledges that "58% of organizations now apply analytics to create a competitive advantage within their markets or industries, up from 37 percent just one year ago" (Kiron *et al.*, 2011). In addition, this study confirms the relevance of business analytics for supporting the development of a customer intimacy strategy, as 62% of the organizations having a strong and sophisticated usage of analytics already leverage analytics for creating personalized relationships with customers. The importance of business analytics is also confirmed in a global survey conducted with 1700 chief marketing officers in

which 81% of the respondents confirmed their intent to use business analytics solutions over the next three to five years (IBM Institute for Business Value, 2011, p.26).

Following a business intelligence approach, the solution *CI Analytics* proposed by this thesis builds upon the idea that customer related data which is stored inside the information system of the provider, such as interactions, activities, projects, and results data contains evidence of customer intimacy. Thus, this thesis aims at finding a set of metrics which can be calculated upon customer related data and which enables the assessment and monitoring of the degree of customer intimacy established by the provider with its customers at both the individual and organizational levels. As suggested by De Choudhury *et al.* (2010), two problems have to be considered: the *inference* and the *relevance* of the customer intimacy metrics. The inference issue relates to the fact that the customer intimacy components are not directly observable and need to be inferred out of existing customer data. For instance, even though previous research indicates a positive association between interactions, customer knowledge, and customer relationships (Ballantyne, 2004), there is no rule establishing that a specific frequency or duration of customer interaction leads to acquiring customer knowledge or establishing customer relationships. The relevance of the customer intimacy metrics is the second main challenge of the customer intimacy assessment proposed by this thesis. An infinite number of metrics can potentially be calculated out of the existing customer data. The challenge is, therefore, to perform the best selection of customer intimacy metrics in order to accurately assess the customer intimacy components. The customer intimacy metrics have to be sorted and weighted according to their relevance for performing this assessment. The *CI Analytics* methodology which is proposed in chapter 5 aims at solving these two challenges of inference and relevance of the customer intimacy metrics.

Focusing on the inference challenge, a central aspect of this thesis and of the solution *CI Analytics* relates to the identification of interaction patterns indicating the development of customer relationships and the acquisition of customer knowledge, and from which some

customer intimacy metrics can be derived. This thesis, thus, intends to provide an innovative contribution to the business analytics subset called *interaction analytics* which has been qualified by Gartner as a technology trigger in the hype cycle for analytics applications (Gartner, 2010). Another key aspect of the solution *CI Analytics* is concerned with the determination of results oriented metrics allowing the assessment of the business impact of the customer intimacy strategy at the organizational level. In that regard, *CI Analytics* relates to the discipline of pattern-based strategies which is defined as the search for "patterns that may have a positive or negative impact on business strategy and operations" (Burton *et al.*, 2011).

In order to perform the customer intimacy assessment, this thesis also relies on network analysis methods (Brandes & Erlebach, 2005b). Such methods have already been successfully applied for assessing relationships in B2B context (Gummesson, 2008, p.296), and previous research already proved that the effectiveness of key account management is affected by the properties of the social network formed by the provider and customer employees, such as the size of the network and the position of the employees in the network (Hutt & Walker, 2006). The solution *CI Analytics* proposed by this thesis therefore provides a graph-based representation of the customer intimacy information and uses network topology metrics such as the degree and closeness centralities as input for the customer intimacy assessment.[1]

The solution *CI Analytics* proposed by this thesis yields the following benefits:

- First, *CI Analytics* provides a systematic graph-based overview of the interactions among provider and customer employees, as well as a visualization of their evolution over time. A change in the interaction pattern can be identified, thereby allowing the provider to proactively act upon it. For instance, frequent interactions with the support team could indicate customer dissatisfaction. A drop in the interaction between two employees could mean that the customer organization has been modified.

---

[1]   Further details are provided in chapter 5.

In both cases, some actions should be taken by the provider on the basis of this information.

- Second, this approach fosters the exchange of customer knowledge among the provider employees. *CI Analytics* enables the identification of the provider employees who own customer related knowledge and who have established relationships with the customer. By making this information available in the form of a graph representation, provider employees having and seeking customer knowledge can identify each other and share tacit customer knowledge. For instance, a service employee who knows the customer could inform his colleague working in sales about the best ways to approach the customer and provide meaningful insights on the customer needs.

- Finally, the solution *CI Analytics* provides the ability to benchmark the effectiveness of the customer intimacy strategy with different customers. The provider can identify to which extent the customer intimacy strategy was executed with each customer and identify which customers are responsive to the customer intimacy strategy. For instance, if the provider invests resources in adapting the solution proposed to the customer, but the customer disregards this solution and selects a cheaper one, then the provider should consider changing its strategy with this customer. *CI Analytics* allows the identification of such patterns out of customer related data.

1.4. Research Questions

The three concrete research questions addressed in the scope of this thesis are derived from the previously outlined research objective.

Research Question 1 – How can the concept of customer intimacy be broken down into multiple assessable customer intimacy components?

Customer intimacy is a complex type of strategy which includes multiple facets. In order to perform the assessment of the degree of

customer intimacy achieved by a provider with its customers, a thorough understanding of the concept of customer intimacy is required. The first research question of this thesis is therefore concerned with the analysis and identification of the key components of the customer intimacy strategy. With the original definition of customer intimacy provided by Treacy & Wiersema (1993) as the starting point of the analysis, this thesis derives multiple measurable and actionable *customer intimacy components*, thereby creating the foundations of the customer intimacy model proposed in chapter 5.

Research Question 2 – Which metrics can be created upon customer related data in order to infer the customer intimacy components?

The second research question of this thesis relates to the conception of *customer intimacy metrics* which can be calculated upon customer data stored in the information system of the provider. These metrics should provide the means to determine the values of the customer intimacy components. This question, thus, relates to the previously introduced challenge of inferencing the customer intimacy components upon customer intimacy metrics. Leveraging literature grounded in the fields of interaction and relationship marketing, some interaction, activity, and result patterns are identified and used to conceive significant customer intimacy metrics. The *CI Analytics* model which is proposed in chapter 5 elaborates on the customer intimacy metrics proposed by this thesis to infer the values of the customer intimacy components. For validation purposes, this model has been embedded in the software *CI Analytics* which was implemented in the scope of this thesis. This software described in chapter 6.

Research Question 3 – Which combination of metrics provides the most accurate assessment of the customer intimacy components?

The third research question is concerned with the determination of the relative importance of the customer intimacy metrics for accurately assessing the values of the customer intimacy components. This question therefore relates to the previously mentioned challenge of relevance of the customer intimacy metrics. Since each provider

manages the relationship with its customers and interacts with the customer employees in a specific way, the relevance of the customer intimacy metrics is influenced by the specific activity and interaction patterns of the provider: some metrics may be relevant for a specific provider and irrelevant for another one. In order to answer this research question, this thesis proposes in chapter 5 the *CI Analytics* methodology for determining the relevance of the customer intimacy metrics and for calibrating them to the activity and interaction patterns of the provider. This methodology is based on data-mining and on machine learning algorithms which are explained in chapter 3. In order to validate the results of this thesis, the *CI Analytics* methodology has been tested in a real-case scenario with the IT software and service provider CAS Software AG.[2] The results of this validation are proposed in chapter 7.

1.5. Structure of the Thesis

As depicted in figure 1.2, this thesis is structured into three parts and eight chapters.

Part 1 presents the foundational and preliminary knowledge which is required in order to understand this thesis.

- Chapter 1 (*Introduction*) defines the context of this thesis, details the research problem, and sets out the research questions addressed by this thesis.

- Chapter 2 (*Towards Customer Intimacy*) elaborates on the concept of customer intimacy as defined in past literature and analyzes its distinctive characteristics with regard to other prominent marketing concepts such as relational marketing, the service-dominant logic, and key account management.

- Chapter 3 (*Methods and Techniques to Assess Customer Intimacy*) lays down the methods and techniques leveraged by this thesis to achieve the objective of assessing customer intimacy upon

---

[2]   Further information on CAS are available at *www.cas.de* (accessed on 10.11.2011).

Part 1 Foundations and Preliminaries	Chapter 1 **Introduction**	
	Chapter 2 **Towards Customer Intimacy**	Chapter 3 **Methods and Techniques to Assess Customer Intimacy**
Part 2 Conceptual Model	Chapter 4 **Customer Intimacy Breakdown Analysis**	
	Chapter 5 ***CI Analytics* Model and Methodology**	
Part 3 Evaluation	Chapter 6 ***CI Analytics* Software**	Chapter 7 ***CI Analytics* Validation**
	Chapter 8 **Conclusion**	

Figure 1.2.: Structure of the Thesis

customer related data available in the provider's information system. More specifically, this chapter introduces fundamental knowledge on graph theory and social network analysis as well as on data mining and machine learning algorithms.

Part 2 elaborates on the conceptual model proposed by this thesis and establishes the means and methodology allowing the assessment of customer intimacy in a B2B context.

- Chapter 4 (*Customer Intimacy Breakdown Analysis*) analyzes the concept of customer intimacy and establishes how it can be broken down in multiple quantifiable components. This chapter, thus, sets out the foundation of the overall model to assessing and monitoring customer intimacy.

- Chapter 5 (*CI Analytics Model and Methodology*) elaborates on the *CI Analytics* model proposed by this thesis to assess customer intimacy and develops a set of metrics allowing the measurement of the customer intimacy components defined in chap-

ter 4. This chapter also develops the *CI Analytics* methodology conceived to use the *CI Analytics* model as well as to calibrate it to the specific patterns of the organization adopting the model.

Part 3 provides an evaluation of the proposed *CI Analytics* model and methodology elaborated in chapters 4 and 5.

- Chapter 6 (*CI Analytics Software*) details the software *CI Analytics* which has been conceived in the scope of this thesis to implement the *CI Analytics* model and to calculate the customer intimacy metrics proposed in chapter 5. This chapter, thus, validates the feasibility of the customer intimacy assessment proposed by this thesis.

- Chapter 7 (*CI Analytics Validation*) develops a real case scenario with the IT software and service provider CAS Software AG in which the *CI Analytics* methodology proposed in chapter 5 has been applied. Following this methodology, this chapter depicts how the customer intimacy metrics have been calculated upon real data related to 14 different customers and calibrated to fit the characteristics of this provider. This chapter subsequently presents the results of the evaluation of this calibration, thereby validating the overall approach of this thesis to assessing customer intimacy.

- Chapter 8 (*Conclusion*) analyzes the extent to which the research questions defined in chapter 1 have been answered by this thesis and summarizes its contribution. This chapter subsequently outlines the managerial implications which can be derived from this thesis, addresses the limitations of its findings, and suggests directions for future research.

2. Towards Customer Intimacy

Several aspects are recurrent when reading literature on customer intimacy such as gaining a competitive advantage, developing a strategy, or managing customer knowledge and relationships. In order to achieve the goal of this thesis to assess customer intimacy, it is therefore necessary to fully understand this concept as well as these different aspects. Moreover, during and after the concept of customer intimacy has been established in 1993, several theories have been proposed which present similarities to customer intimacy.

This chapter will elaborate on the value discipline customer intimacy and analyze its distinctive characteristics. Section 2.1 will develop customer intimacy according to its original definition proposed by Treacy & Wiersema (1993, 1997), and highlight its differences to two other value disciplines, namely product leadership and operational excellence. Section 2.2 will establish why customer intimacy is anchored in the theory of relationship marketing and is related to the modern perspective on services called the service-dominant logic. Finally, section 2.3 will outline more specifically three approaches related to the implementation of customer intimacy, namely key account management, market orientation, and customer relationship management.

2.1. Three Value Disciplines to Achieve Market Leadership

The concept of customer intimacy was first introduced by Treacy & Wiersema (1993) and emerged from a systematic three-year analysis of 80 corporations acting worldwide in 40 different business to business (B2B) and business to consumer (B2C) markets. The goal of this research was to identify patterns among organizations that were market leaders. Treacy and Wiersema established objective criteria to explain the reasons for the success or failure of these firms. They were able to thereby uncover previously hidden sources of competitive advantage for companies operating in these markets. In order to achieve this objective, they analyzed the different facets of these organizations. First, they considered the value propositions of these firms, looking at the implicit promises made to their customers. These value propositions include multiple factors such as price, performance, quality, and scope of the offering. They subsequently analyzed the operating models of the organizations, which include all the components required to deliver value to the customer, such as business processes, business structures, management systems, culture, and information technology. Finally, they introduced the novel concept of *value disciplines*. Value disciplines are types of strategy on which the strategy is aligned and are accordingly different from strategy or strategic goals.

In their study, Treacy and Wiersema identified three distinct value disciplines: operational excellence, product leadership, and customer intimacy. As depicted in figure 2.1, operational excellence refers to a focus on delivering the highest price-quality ratio, or so called *best total cost* for the customer. Product leadership concerns organizations that provide their customers with the highest quality and most advanced innovations. It can be summarized as *best product*. Finally, customer intimacy driven organizations may not deliver the cheapest solution nor the latest innovations to the market, but instead of focusing on average market requirements, they have the ability to develop individualized solutions, tailored to the exact needs of each of their customers. This value discipline can be understood as providing

customers with the *best total solution*. These three value disciplines shares similarities with the generic competitive strategies proposed by Porter (2004). The value discipline operational excellence is close to the generic competitive strategy "overall cost leadership". Product leadership resembles the strategy "differentiation" and customer intimacy shows commonalities with the strategy "focus".

Product Leadership
"Best Product"

Product
Differentiation

Operational
Competence

Customer
Responsiveness

Operational Excellence
"Best Total Cost"

Customer Intimacy
"Best Total Solution"
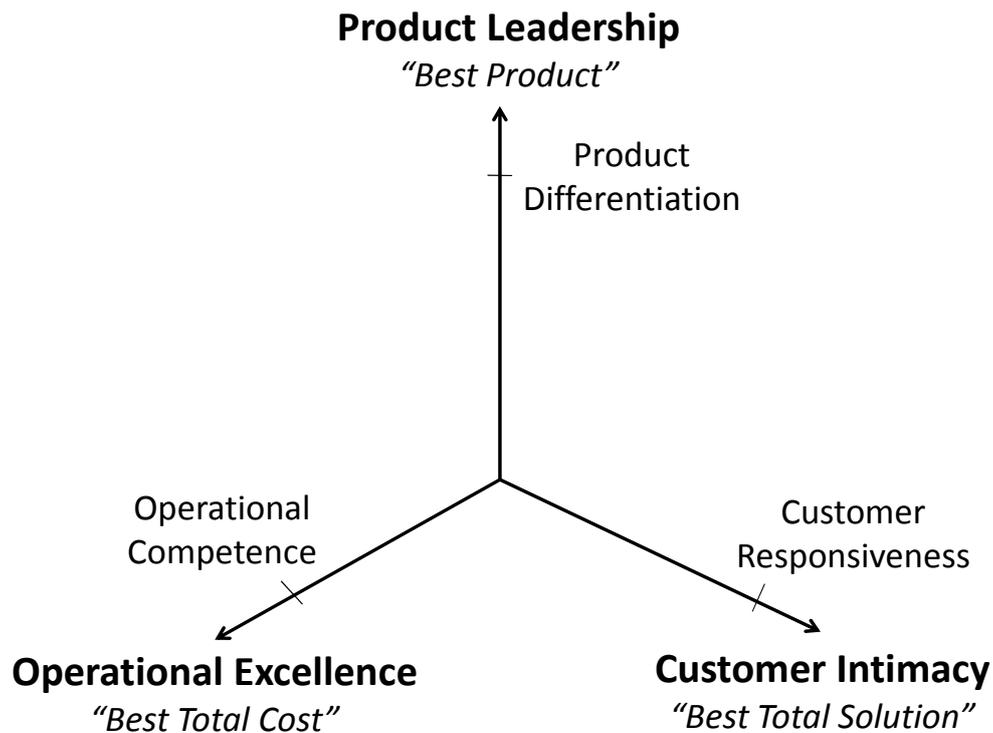
Figure 2.1.: Three Value Disciplines to Achieve Market Leadership (Treacy & Wiersema, 1997, p.45)

The central argument of Treacy and Wiersema's thesis is that, in order to become a market leader, an organization should choose one and only one value discipline and align the two other facets accordingly, that is the value proposition and the operating model of the organization. The chosen value discipline is the one by which the organization will gain its market reputation and achieve clarity in the perception of its customers. It determines all subsequent business decisions related to the value proposition and to the operating model.

Selecting one specific value discipline, however, does not mean that the other two value disciplines become irrelevant for the organization. Treacy and Wiersema nuance their argumentation by stating that organizations should strive for excellence in one of the three value disciplines, and achieve a certain threshold in the other two. Indeed, an organization following operational excellence will not be successful if its products or services do not achieve a certain degree of quality. A customer intimacy driven company must keep its prices within reasonable limits for the customers and deliver high quality solutions. However, instead of being the first one on the market to propose a new feature, this company will work closely with its customers, evaluate to which extent and in which form they need this feature, and deliver it later than product leadership driven firms would, but in a way that fits its customers' requirements. Treacy and Wiersema argue that organizations striving for excellence in all three value disciplines do not achieve to become market leaders. As a matter of fact, this lack of commitment to one value discipline leads to a dilemma for every business decision taken in the organization. The trade-off between creating the highest quality product and delivering with the cheapest costs is not performed in a consistent way, leading to hybrid value propositions which are ambiguous in the eyes of the customers. In a similar way, the choice to propose standard market offerings rather than to design solutions for individual customers might be questioned every time a customer raises a new requirement. This increases complexity significantly, causes confusion in the organization management and leads to "doing business with yourself rather than with your customers" (Treacy & Wiersema, 1997, p.45).

The next part of this section briefly summarize the value disciplines operational excellence and product leadership in order to contrast them in section 2.1.2 with the value discipline customer intimacy. The core value proposition as well as the operating model are described for each of the three value disciplines as they were originally presented by Treacy & Wiersema (1997).

2.1.1. Operational Excellence and Product Leadership as Alternatives to Customer Intimacy

2.1.1.1. The Value Discipline of Operational Excellence

The value proposition of operationally excellent organizations was labeled by Treacy & Wiersema (1993) as providing the best total cost to the customer. They consciously used the word cost instead of price, as the objective is to lower the overall costs incurred to the customer with the purchased product or service. This includes the price paid by the customer, but also additional factors such as the time spent by the customer to purchase the product or to obtain support. Such companies tend to offer the best price quality ratio on a limited and precisely defined set of products or services. Low cost airline companies such as South-West or EasyJet are classic examples for operational excellence. The promise to the customer is limited to transporting the customer from departure to destination in a certain amount of time and very few additional options are available for free to the customer: the booking and payment must be performed via internet, there is only one comfort category, and no extra services are provided during the flight. These enterprises present a clear value proposition: their customers are aware that they should not expect anything beyond the standard offering, neither to hope for rewards for their loyalty, but they also know that the price for these standard offerings is unmatched by other companies.

Standardization, norms, and procedures are at the heart of the operational excellence operating model. Operational excellence driven organizations eliminate defects and remove variation in order to lower costs and to guarantee high quality levels. In this regard, driving operational excellence shares commonalities with the adoption of lean management and six sigma programs.[1] Operationally excellent

---

[1] Lean management aims to accelerate the velocity of any process by reducing waste in all its forms. Six sigma is a set of practices originally developed by Motorola to systematically improve processes by eliminating defects. Six sigma uses rigorous data analysis to pinpoint the source of errors that contribute to process variations (George, 2003).

companies thoroughly standardize their business processes throughout the entire organization and try to include their providers' activities in these processes. Indeed, vertical integration and tight partnerships with business partners allow them to reduce intermediary costs such as communication, inventory, and transportation costs.

From a cultural standpoint, the operationally excellent organization rewards efficiency. Employees are given a set of standard tasks to accomplish along specific procedures and are expected to complete them with no variations from the rules. They are rewarded when they demonstrate a dedication to fulfill the promises made to the customer, rather than when they show creativity or originality. According to Weill & Ross (2004), firms that pursue operational excellence make large investments in IT systems in which they recognize the ability to lower costs and, thus, to increase their competitiveness. They focus on business process management systems[2] to centralize the coordination and control of the activities and to automate routine and non value adding tasks. They also leverage systems that enable them to automate transactions and facilitate communication with both customers and providers.[3]

2.1.1.2. The Value Discipline of Product Leadership

Companies that pursue the product leadership value discipline intend to deliver the best product or service to the market, in terms of performance and quality, but also with regard to the degree of innovation of their offering. This leads to a clear value proposition, which target customers who have the willingness to pay a premium fee for outstanding quality and who value originality and exclusivity of new features. In many cases, such companies manage to establish an emotional connection to their customers via the provided products and services. Stern (1997) demonstrated that B2C relationships are more

---

[2] Elzinga *et al.* (1995, p.119) define business process management as a "systematic, structured approach to analyze, improve, control, and manage processes with the aim of improving the quality of product and services. BPM is the method by which an enterprise quality program is carried out."

[3] See for instance the EDIFACT standard for inter enterprise data exchange *http://www.unece.org/trade/untdid/welcome.htm* (accessed on 10.11.2011).

likely to be emotional, while B2B relationships are grounded on a rational basis. Many product leadership driven organizations therefore predominantly target B2C markets. Apple[4] is today's typical example for product leadership: Apple delivers a clear message, arguing they provide the highest quality and the most innovative phones and computers. The market recognizes that they keep this promise and hails the fact that their products are almost always radically different from, and significantly better than, those of their competitors. Even though Apple has also reached certain thresholds in operational excellence and customer intimacy, this company has achieved its reputation by differentiating its products from more standard market offerings.

In contrast to operationally excellent organizations, which are driven by procedures and a thorough attention to maintain costs at a low level, the operating model of product leadership companies emphasizes talents of their employees for generating, promoting, and implementing new ideas. Product leaders rely on their ability to innovate and to bring new forms of value to their customers. They need to develop management and organization systems that reward creativity and foster collaboration among the employees. Therefore, experimentation and risk are key aspects of their culture: employees are given the time and resources to try new ideas, and to validate their potential value on the market. Product leaders tend to be organized in small business units having high degrees of autonomy. This form of structure, described as a federal model by Weill & Ross (2004), promotes the development of an entrepreneurial and risk prone environment, as it simplifies and speeds up decision-making processes.

The core business processes of such organizations are two-fold. On the one hand, internally, they design business processes that encourage the diffusion of knowledge and expertise to promote new ideas. These processes support the coordination of the various activities,

---

[4]   Apple Inc. reported its best results ever in year 2010 with 71% revenues growth and 78% earnings growth. Further details are available at *http://www.apple.com/pr/library/2011/01/18results.html* (accessed on 17.10.2011).

but not in the sense operationally excellent organizations perform it, with strict and rigorous control: they preserve a certain degree of freedom and autonomy for employees and teams. On the other hand, focusing on the external perspective, the business model of product leaders remains sustainable only if the company is able to anticipate the needs of the market and brings its offering to the market before competitors do.[5] Therefore, the second part of their core processes seeks fast commercialization and market exploitation. They design processes that allow them to reduce time to market, by speeding up engineering, production and delivery phases. Product leaders also have effective communication campaigns for their new product or services launch and well established marketing plans. Indeed, their customers must acknowledge the superior value of their offering in order to accept the payment of a premium fee. Thus, they need to be able to clearly articulate the benefits of their value proposition to prepare the market, and to develop a demand for products and features that did not exist in the past.

2.1.2. The Value Discipline Customer Intimacy

This section outlines the characteristics of the value discipline customer intimacy, as first introduced by Treacy & Wiersema (1993, 1997), and contrasts them with those of the previously described value disciplines operational excellence and product leadership. While operationally excellent companies focus on lowering total costs and product leadership firms try to bring the best products on the market, customer intimacy driven organizations aim at providing each of their customers with the best solution.

2.1.2.1. Value Proposition

The uniqueness of customer intimate organizations is that, instead of focusing on the market and trying to fulfill the most demanded market requirements, they are able to focus specifically on each of their

---

[5]   For instance, the Apple iPad arrived on the market in April 2010, a full year
       earlier than similar products from competition.

customers and their individual needs, problems, expectations. They demonstrate to their customers a clear value proposition which goes beyond mere delivery of products and services: customer intimate organizations apply their knowledge to investigate the customer's specific problems, in cooperation with the customer employees, and design solutions that include customized versions of the products and services they intend to sell. Then, they actively control the deployment of the solution in order to ensure that the customer's expectations are actually reached. Customers see such companies as trusted partners on which they can rely upon. It may be that other alternatives on the market are cheaper or more innovative, but a customer intimate organization brings the confidence to its customers that its solution is solid, tested, and actually delivers the expected benefits. In fact, customer intimate organizations demonstrate their commitment to their customers by assuring that their solutions will deliver the promised results in a mutually beneficial manner: while customers focus on the part of their operating models that are critical to their own success, the customer intimate firm manages their secondary processes. For instance, the customer intimate firm will take over the IT organization of its customers in the form of an outsourcing contract which guarantees that certain levels of performance, quality and flexibility are achieved.

In order to develop a sustainable competitive advantage, customer intimacy driven organizations not only fulfill the needs of their customers, but they anticipate the customer problems and identify sources of value for their customers in order to create some demand in the customer organization. Therefore, companies that pursue a customer intimacy strategy heavily rely on their insight of the customer industry, on the customer related knowledge they acquired, and on the interpersonal relationship they developed inside the customer organization. In that regard, Abraham (2006, p.1) complements the original definition of customer intimacy and states that customer intimacy is concerned with "the formal or informal set of relationships established between suppliers and customers, with a diverse array of partners, from corporate leadership to functional leadership (engineering, marketing, operations, maintenance, or service) and end-

users of products or services". While operationally excellent organizations benefit from their optimized processes and product leaders take advantage of their innovations, customer intimates firms main asset is the loyalty of their customers (Treacy & Wiersema, 1997).

2.1.2.2. Operating Model

Figure 2.2, originally presented in Treacy & Wiersema (1997, p.130), depicts the customer intimacy operating model. In order to deliver tailored solutions to their customers, customer intimate companies need to establish an operating model that allow them to provide a broad and deep level of support and services to their customers. This means that all entities of the enterprise, sales and services, but also product development and manufacturing must be oriented towards the objective to satisfy the needs of the customer. If a sales representative thoroughly understands the needs of the customer, but he is restrained to provide an adequate solution because his organization does not adapt a product or a service, then this enterprise does not achieve customer intimacy. Treacy & Wiersema (1997, p.133) call this "customer responsiveness", carefully understanding the customer needs, showing empathy for the customer problems, but not being able to provide a satisfactory offering to the customer. Batt (2004, p.172) confirm that "the firm must keep deepening its knowledge of the customers and put this knowledge to work through the organization." Consequently, customer intimate companies must empower the front line employees that have understood the customer requirements and provide them with the means to leverage the skills and capabilities of the entire organization to build the solution. Such companies, therefore, emphasize structures aligned with the customer base and the development of decentralized entrepreneurial account teams, who take responsibility for budgets, prices, technological choices, and communication.

Moreover, an organization can rarely deliver a total solution to its customers solely out of its own assets. The wide variety of customer needs and requirements lead customer intimate companies to develop partnerships with subcontractors. Treacy & Wiersema
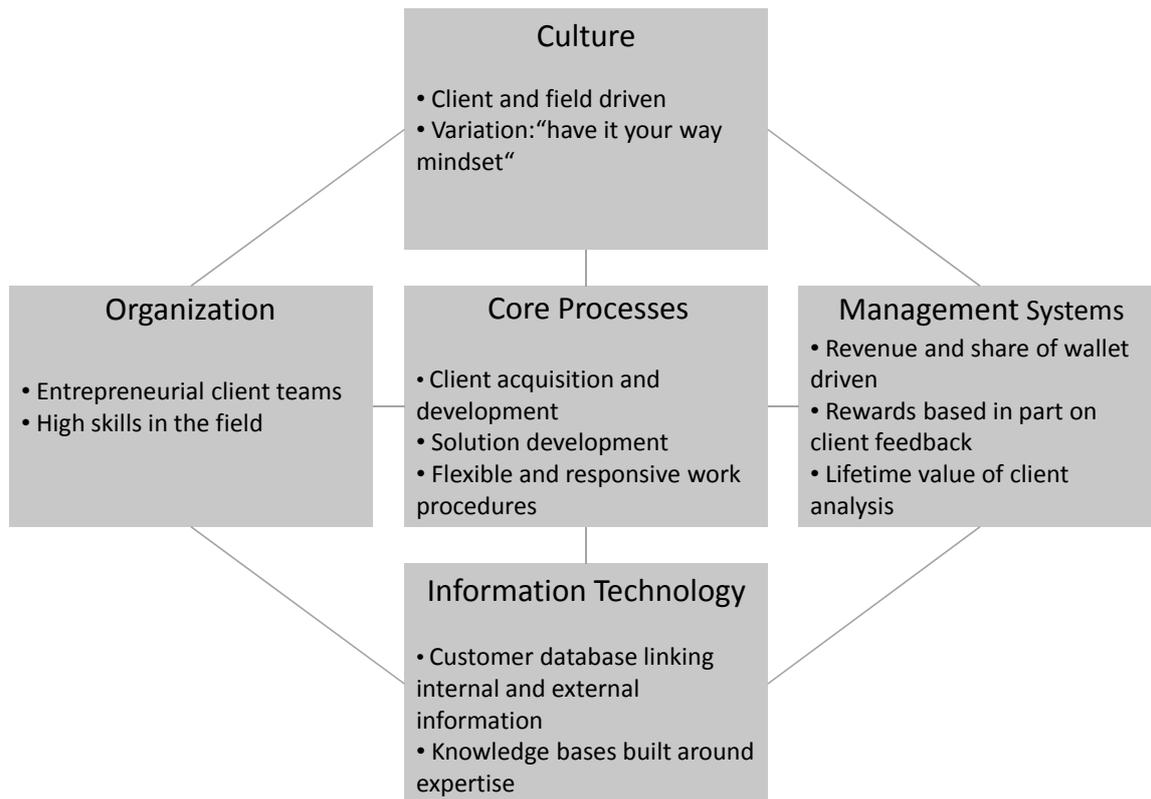
Figure 2.2.: Customer Intimacy Operating Model (Treacy & Wiersema, 1997, p.130)

(1993, p.137) argue that customer intimate organizations are built upon "hollow delivery systems" and their strength "lies not in what they own, but in what they know and how they coordinate expertise to deliver solutions." Customer intimate organizations tend to take the role of a resource integrator between the customer and a large range of operationally excellence and product leadership driven contractors to ensure that the customer receives the best offer, in terms of features, price, and quality. Consequently, the core processes of customer intimate organizations should be based on flexible and solution-driven work procedures that facilitate collaboration inside the organization as well as with business partners.

Not all customers are responsive to a customer intimacy driven strategy. Several enterprises will take their decision based on price or product features and, thus, are reluctant to pay a premium fee for the

value of the expertise provided by customer intimate firms. Organizations which are most inclined to partner with a customer intimate company exhibit some specific characteristics with regards to their attitude, operational fit, and financial potential (Treacy & Wiersema, 1997, p.139). The attitude refers to the willingness of the customer to engage in a business relationship. Indeed a relationship exists only if the customer perceives a mutual benefit, an opportunity for an ongoing association, and if he has the readiness to loose some independence in return (Donaldson & O'Toole, 2007, p.58). A customer that does not demonstrate a certain degree of loyalty should not remain a target of firms pursuing a customer intimacy strategy. The second characteristic refers to the operational fit. This operational fit exists if the provider's expertise matches a deficit of competence on the customer side. Indeed, if the customer recognizes the superior skills of the provider, he will rely on him to provide the overall solution. However, if the customer is already too knowledgeable in the concerned area, he may favor another offering and find it himself on the market. Since most organizations have developed an expertise in their core business, customer intimate organizations mainly look for this expertise gap in the supporting functions of the organization, such as information and communication technology, finance, or communication in order to create this operational fit. The last characteristic relates to the customer's financial potential. This financial potential should be large enough as well as distributed on a long-term period of time for the customer to be a target of the customer intimate organization. Customer intimate companies invest significant amounts of time and resources to gather and manage customer related knowledge as well as to generate knowledge that is relevant for the customer, such as insight in his industry. The return on this investment is derived from long-term regular cash-flows from the customer, and short-term single transactions are not profitable for organizations that base their business model on customer intimacy (Treacy & Wiersema, 1997, p.140).

As a result, the management system of customer intimate organizations should support the identification and acquisition of customers presenting such characteristics as well as to help retaining them. Its

key performance indicators are not related to market shares, but to account penetration, shares of customers' spendings, and customer retention. Similarly, the sales force is driven by two objectives: they should acquire new customers and they should provide an ongoing support to their existing customers. The network of interpersonal relationships established between the employees of the customer intimate firm and the employees of its customers is, therefore, a key factor of success (Gummesson, 2008, p.91). From a technological perspective, as presented in the introduction, customer intimate organizations develop customer relationship management systems that allow them to achieve a single view of the customers, as well as knowledge database to foster the dissemination of customer knowledge.

The next section sets out the similarities between customer intimacy and the established concepts of relationship marketing, service marketing, and service-dominant logic.

2.2. Customer Intimacy: Grounded in Relationships and Services

This section establishes the relation between customer intimacy and the notions of relationship marketing, service marketing, and the modern view on services, namely the service-dominant logic. Section 2.2.1 introduces the notion of relationship marketing and contrasts it with the transactional perspective on marketing. Then, section 2.2.2 outlines the importance of services for the development of relationship marketing and elaborates on a relationship marketing perspective which is particularly important in the scope of this thesis, namely the "Nordic School of Thought" (Gummesson, 1996). In section 2.2.3, the service-dominant logic is presented as the evolution of relationship marketing. Finally, in order to fully understand the concept of customer intimacy, section 2.2.4 contrasts customer intimacy with relationship marketing and with the service-dominant logic.

2.2.1. Two Divergent Perspectives on Marketing

While the notion of marketing as a distinct discipline arose in the beginning of the 20th century, the emphasis on relationships has only received attention over the past 40 years (Sheth & Parvatiyar, 2000). Arndt (1979), introducing the concept of "domesticated markets", was one of the first to establish the importance of developing long-lasting relationships with key customers rather than focusing on single transactions. Adler (1966) and later Varadarajan & Rajaratnam (1986) outlined the widespread acceptance of symbiotic marketing as a way to achieve sales and profit growth, with an emphasis on collaboration and strategic partnership for mutual benefit of the parties. The first formalization of the concept of relationship marketing in literature occurred in 1983: Berry (1983, p.25) defined relationship marketing as "attracting, maintaining and – in multi-service organizations – enhancing customer relationships." Since then, multiple perspectives have emerged with an emphasis on a variety of themes such as quality, customer service, alliance and partnerships, communication and interaction (Mohr & Nevin, 1990; Christopher *et al.,* 1993; Morgan & Hunt, 1994; Varadarajan & Cunningham, 1995).

The first part of this section focuses on the exchange perspective and defines transactional marketing. Then, the second part elaborates on the relationship perspective and defines relationship marketing in contrast to transactional marketing.

2.2.1.1. Transactional Marketing

The concept of transactional marketing originated in the industrial era, as a consequence of mass production, mass consumption, and the division of labor. In these provider driven markets, the most important challenge was to optimize production capabilities and employees productivity in order to increase the produced volumes, thereby achieving economies of scale. Low priced and standardized products replaced customized offerings. New specialized organization structures with dedicated purchasing and selling functions fundamentally changed the way of doing business. Sheth & Parvatiyar (2000) argue

that the providers and customers have been separated. Business relationships have become impersonal or even replaced by intermediaries,[6] such as wholesales companies and distributors, whose roles, acting as agents, are two-fold: first, they have to handle the stocks produced and, second, they have to distribute these goods into the market.

The exchange perspective on marketing arose in the early 1960s when most markets became saturated and competition increased (Wengler, 2006). New marketing practices emerged, "focused on sales, advertising and promotion, for the purpose of creating new demand to absorb the oversupply of goods" (Sheth & Parvatiyar, 2000, p.130). Marketing functions were implemented as a means to locate and persuade potential customers to purchase more goods and services in order to increase sales volumes and generate additional profits (Gruen, 1997). In this perspective, customers are not considered as single and *active* entities, but aggregated in *passive* market segments. As depicted in figure 2.3, the focus is on the outcome of the transactions: marketing aims solely at supporting sales activities, rather than on developing and maintaining business relationships. Value creation and value distribution are two distinct activities, and marketing concentrates on the latter one only: the customer is solely considered as the receiver of value distributed by the firms, and does not participate in the creation of value.

2.2.1.2. Relationship Marketing

In the 1980s, as customers expectations were raising, companies started to search for new means of generating value. The exchange perspective focused on single transactions was questioned and new programs dedicated to the partnership with individual customers emerged (Shapiro & Wyman, 1981). Later, several publications recognized the potential of collaboration and cooperation between buyers and sellers to develop a competitive advantage (Narver & Slater, 1990; Varadarajan & Rajaratnam, 1986). The findings from Reichheld &

---

[6] These intermediaries are called "middlemen" in Sheth & Parvatiyar (2000).

Process

Relationship
Perspective

Value Distribution   Value Creation

Exchange
Perspective

Outcome
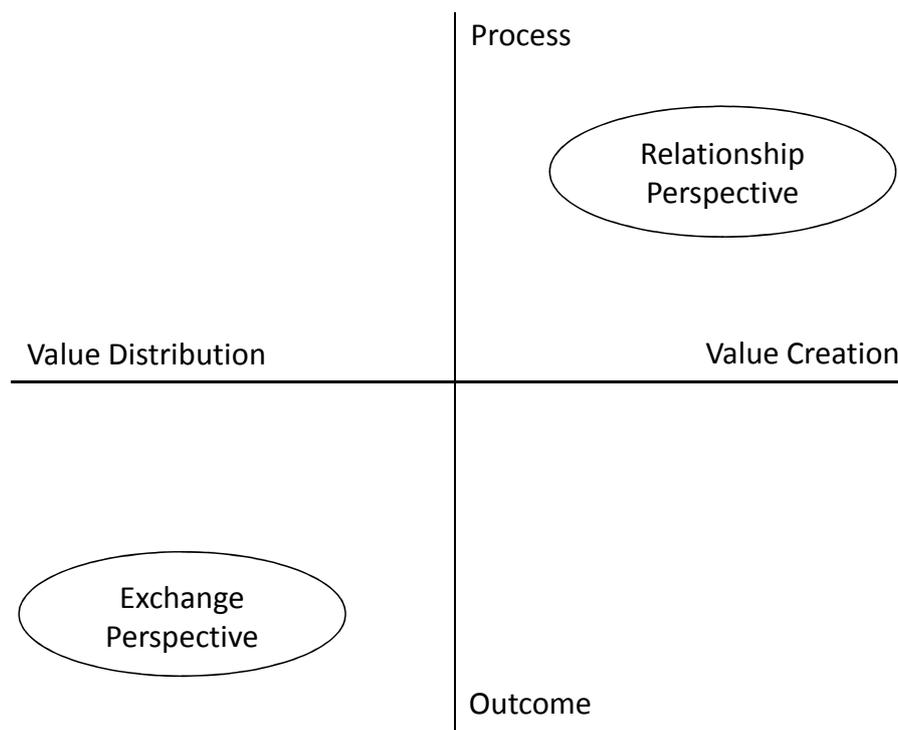
Figure 2.3.: Exchange and Relationship Perspectives (Sheth & Parvatiyar, 2000)

Sasser (1990) that a 5% improvement in customer retention can result in a profitability improvement comprised between 25% and 85% created a strong impulse for research that investigates the association between customer loyalty, retention, and satisfaction (Dick & Basu, 1994). Later, various studies argued that customer satisfaction could be better achieved through an emphasis on customer relationships, with the objective to retain valuable customers, rather than through a focus on single transactions (Day & Montgomery, 1999). Moreover, the development of new technological solutions and the growth of the service economy changed the market dynamics and boosted the development of the relationship marketing concept. Indeed, new IT based solutions and the rise of internet services enable selling and buying firms to reestablish direct contact, without the needs of in-

termediaries.[7] The growth of the service economy and the ongoing shift to servitized businesses further reduce the needs for intermediates, as services are often directly provided by the provider. Indeed, as developed in section 2.2.2, services enable the development of relationships and much literature on service marketing is devoted to relationship marketing (Grönroos, 2007; Lovelock & Wirtz, 2007).

As illustrated in figure 2.3, Sheth & Parvatiyar (2000) defined the two axes of the relationship perspective in contrast to the exchange perspective. The first dimension outlines that value is not only distributed to the customer, but created in collaboration with the customer. Higher value can be generated if the customer actively cooperates with the provider and shares his knowledge and expertise. Marketing, thus, should not unilaterally focus on convincing customers of the benefits of the value proposition, but it should involve the customers in the definition and development of a joint value proposition which is mutually beneficial for both parties. The second dimension refers to the process dimension of relationship marketing. Relationship marketing requires to establish a set of processes focused on the initiation, maintenance, and termination of business relationships, rather than on the outcomes on the relationship. In line with these two dimensions, Parvatiyar & Sheth (2000, p.9) propose to define relationship marketing as "the ongoing process of engaging in cooperative and collaborative activities and programs with immediate and end-user customers to create or enhance mutual economic value, at reduced cost."

The next part of this section emphasizes the importance of services for the development of relationship marketing and associates service marketing to relationship marketing.

2.2.2. The Service Dimension of Relationship Marketing

Services in contemporary times have become the most important driver of Western economies as they represent over 70% of the Gross

---

[7]   Parvatiyar & Sheth (2000) call this transformation the deintermediation process by which producers and customers directly interact with each other.

Domestic Product (GDP) in both Europe[8] and United States.[9] However, it remains challenging to characterize services as they refer to a wide variety of concepts, and were described differently in various disciplines such as information technology, service design, or marketing. Focusing on business services and on the marketing perspective, Grönroos (2007, p.25) states that a service is "a process consisting of a series of more or less intangible activities that normally, but not necessarily always, take place in interactions between the customer and service employees and/or physical resources or goods and/or systems of the service provider, which are provided as solutions to customer problems." In this definition, three aspects characterizing services are important: customer interaction, service intangibility, and service individualization. Bruhn & Georgi (2006) confirm the relevance of these characteristics in their three dimensional continuum along which both products and services are positioned: the three dimensions of this continuum are interactivity, intangibility, and individuality. They argue the more an offering is interactive, intangible, and individualized, the more this offering is considered as a service. A detailed analysis of these characteristics establishes the reasons why relationships are embedded in services and why service marketing closely relates to relationship marketing:

- *Customer interaction* refers to the involvement of the customer in the process of delivering the service. Several activities of this process include the customer employees and provide them with the opportunity to communicate, exchange information and knowledge and, thus, to participate to some extent to the creation of value, together with the provider employees. The value of a service is not consumed as an outcome by the customer at the end of the service process as a product would be, but simultaneously during the service process. Consequently, services are by definition aligned to the two dimensions of the previously presented relationship perspective: a focus on the

---

[8]   See *https://www.cia.gov/library/publications/the-world-factbook/geos/ee.html* (accessed on 10.11.2011).

[9]   See *https://www.cia.gov/library/publications/the-world-factbook/geos/us.html* (accessed on 10.11.2011).

value creation process rather than on the value creation out-
come, and a value which is created *with* the customer rather
than distributed *to* the customer.

- *Intangibility* refers to the fact that a service do not systemati-
cally result in a tangible outcome. Travel or hotel services for
instance are intangible: customers who purchase a service ex-
perience are left without any "tangible good" at the end of the
service delivery. More specifically, customers cannot see and
test the service characteristics prior to purchasing it, as they
would with a product. They have to trust the provider in its
ability to deliver the agreed service, and to demonstrate a will-
ingness to establish business relationships with their trusted
service partners. This fact is confirmed by several studies which
acknowledge the positive association between trust and rela-
tionship in a service context (Palmatier *et al.*, 2006; Berry, 1995;
Morgan & Hunt, 1994).

- *Individualization* concerns the ability of the provider to customize
his offering in order to fulfill the customer requirements. This
aspect refers to the strategic perspective on relationship marke-
ting. Indeed, the concept of relationship marketing emerged
as companies were seeking new sources of competitive advan-
tage and new means to generate value. Considering each cus-
tomer on an individual base rather than focusing on market
segments is the essence of relationship marketing and, there-
fore, the servitization of the offering is the means to adopt a
relationship marketing strategy (Berry, 1983, p.26).[10]

The "Nordic School of Thought", originating in Sweden and Finland,
is recognized as the pioneer and leader in service marketing. It has
established a direct association between service marketing and rela-
tionship marketing (Gummesson, 1996). This approach is led by two

---

[10] Berry (1983) establishes the five strategy elements for practicing relationship
marketing: "developing a core service around which to build a customer
relationship, customizing the relationship to the individual customers,
augmenting the core services with extra benefits, [...] pricing services to
encourage customer loyalty, and [...] marketing to employees so that they, in
turn, will perform well for customers."

prominent researchers in the field of relationship marketing, Grönroos and Gummesson, who elaborated two definitions of relationship marketing:

- Grönroos (1997, p.407) presents a definition which is close to the original one proposed by Berry (1983), and emphasizes the notion of relationship process and the development of a partnership to achieve the objectives of both parties. He states that relationship marketing is "the process of identifying and establishing, maintaining, enhancing, and when necessary terminating relationships with customers and other stakeholders, at a profit, so that the objectives of all parties involved are met, where this is done by a mutual giving and fulfillment of promises."

- Gummesson (1995)[11] proposes a definition that emphasizes the notion of interactions. He argues that relationship marketing is "marketing seen as interactions, relationships, and networks," where "relationships are contacts between two or more people, but they also exist between people and objects, symbols and organizations". He also defines networks as "sets of relationships", and interactions as "activities performed within relationships and networks" (Gummesson, 1996, p.33). This perspective is particularly relevant for this contribution, as the model proposed in chapter 5 is based on an analysis of interaction data.

The next part of this section introduces the service-dominant logic as an evolution of the concept of relationship marketing.

2.2.3. The Service-Dominant Logic as an Evolution of Relationship Marketing

In 2004, the prestigious Journal of Marketing published an article entitled "Evolving to a new dominant logic for marketing" (Vargo & Lusch, 2004a) which brings the concepts of relationship marketing

---

[11]   This citation is presented in Gummesson (1996).

and service even closer. This article includes controversial ideas that were discussed at length before and after its publication: it was accepted for publication only after a five year review process (Bolton, 2006). The authors claim that "marketing has moved from a goods-dominant view, in which tangible output and discrete transactions were central, to a service-dominant view, in which intangibility, exchange processes, and relationships are central" (Vargo & Lusch, 2004a, p.2). They also bring accordingly a new perspective on the service concept which they define as "the application of specialized competences (knowledge and skills) through deeds, processes, and performances for the benefit of another entity or the entity itself" (Vargo & Lusch, 2004a, p.2).

The authors' argumentation for a new logic of marketing consists of multiple foundational premises which are thoroughly described with reference to past literature mainly rooted in the field of relationship marketing (Vargo & Lusch, 2008a, p.7). The eight original foundational premises are italicized in this and the next paragraphs. First, in contrast to many theories that differentiate tangible goods on one side and services on the other side, the thesis of Vargo and Lusch unifies goods and services: goods exchanged within economic transactions are actually the "outcome" of services understood as application of the provider's knowledge and skills. Thus, even though goods are traded between economic actors, *service is the fundamental basis of exchange.*" Instead of separating goods from services, the service-dominant logic distinguishes two types of resources: the operant and the operand resources. Operant resources are active and possess knowledge and skills, which have to be applied on other resources in order to generate value. For instance, if someone has some special knowledge or skills, but does not utilize them, then no value is created. On the other hand, the operand resources, such as goods and natural resources, are static and inherently contain the outcomes of the application of the operant resources knowledge and skills: the operant resources perform transformation actions on the operand resources in order to increase their potential value, once the customer utilize them. Therefore, goods, in this logic, embed the knowledge and skills of the provider: *goods are distribution mechanisms for service*

provision".

As described in section 2.2.1.1, the emergence of intermediaries in the 20th century, such as distributors and wholesales companies, has led to separate providers and customers who are no longer in direct contact: providers trade with intermediaries and intermediaries trade with customers. It has resulted, on the long term, in hiding that the application of knowledge and skills is the essence of economic transactions: *"indirect exchange masks the fundamental basis of exchange."* Moreover, Vargo and Lusch argue that the dominance of the manufacturing perspective and the segmentation of the economy into eras, such as the agricultural and later the industrial era, have focused the analysis of economic activities on the optimization of goods production efficiency. As a result, the tangibility dimension has received an overly important consideration and intangible items have been simply perceived as side elements. In contrast to this perspective, the service-dominant logic states that intangibility is only one aspect that characterizes economic exchanges and, therefore, *"all economies are service economies."* Since all economic exchanges are derived from the application of operant resources such as knowledge and skills, the service-dominant logic states that *"operant resources are the fundamental source of competitive advantage."* Indeed, the added value that leads to a competitive advantage may lie in the characteristics of the product or service sold to the customer, but primarily results from the leverage of knowledge and skills to fulfill the customer's needs and solve his problem.

The remaining characteristics of the service-dominant logic outline its association with relationship marketing. Vargo and Lush argue that value, as for relationship marketing, is created together with the customer rather than distributed to the customer: *"the customer is always a co-creator of value."* They emphasize the role of the customer as an operant resource that uses its skills and knowledge as well as the significance of the interaction with the customer in order to increase the created value. Then, the service-dominant logic also differentiates the value-in-exchange from the value-in-use and *"the enterprise cannot deliver value, but only offer value propositions."* Indeed, even in the case of manufactured goods, the actual value is only created

once the customer is using the good. As long as this is not the case, the provider has only proposed value to the customer. Finally, the authors state that *"a service-centered view is customer oriented and relational,"* and describe in four arguments this service-centered view: "(1) identify or develop core competences, the fundamental knowledge and skills of an economic entity that represent potential competitive advantage; (2) identify other entities (potential customers) that could benefit from these competences; (3) cultivate relationships that involve the customers in developing customized, competitively compelling value propositions to meet specific needs; (4) gauge market place feedback by analyzing financial performance from exchange to learn how to improve the firm's offering to customers and improve firm performance" (Vargo & Lusch, 2004a, p.5).

In summary, the service-dominant logic is closely aligned with the relationship perspective (Vargo & Lusch, 2006) and shares its two dimensions "value creation" and "process".[12] First, in both approaches the value is co-created by the provider and the customer rather than created by the provider and distributed to the customer: the customer is an active participant, an operant resource. Second, the focus is on the activities that lead to value creation rather than on the outcome. These activities, called the value creation process in relationship marketing are defined in Vargo & Lusch (2004a, p.2) as "the application of specialized competences (knowledge and skills) through deeds, processes and performances" and represent the substance of economic exchanges in the service-dominant logic. Importantly, the service-dominant logic strengthen the significance of knowledge and its application in order to individualize the value proposition and achieve a competitive advantage even more than relationship marketing: knowledge has already established as an important dimension in relationship marketing, but it is, in the service-dominant logic,

---

[12]   The authors, however, do not oppose the service-dominant logic to the exchange perspective. As they compare the service-dominant logic with the exchange perspective, Vargo & Lusch (2006, p.48) consider that the service-dominant logic "bridges the exchange and relationship perspective and, therefore, obviates the apparent need for abandoning the exchange paradigm."

defined as the fundamental source of competitive advantage.

The next part of this chapter establishes why customer intimacy is a value discipline grounded in the concepts of relationship marketing and service-dominant logic.

2.2.4. Customer Intimacy: A Relationship and Service Based Value Discipline

It is possible to establish an association between customer intimacy, relationship marketing, and the service-dominant logic by comparing the previously introduced descriptions of these three notions. In contrast to the value disciplines product leadership and operational excellence, customer intimacy is rooted in the concept of relationship marketing and it shares commonalities with the service-dominant logic. This statement is motivated by the following four arguments which are elaborated in the next paragraphs: similarly to relationship marketing and the service-dominant logic, (i) customer intimacy supports the idea that value is co-created by the provider and the customer; (ii) customer intimacy focuses on relationship processes established with customers rather than on the delivery of produced outcomes; (iii) customer intimacy does not specifically distinguish tangible products and intangible services; and (iv) customer intimacy recognizes that knowledge is the main source of competitive advantage.

- **The provider and the customer co-create value**
  The first dimension of the relationship perspective which is described in section 2.2.1.2 states that value is created together with each customer rather than produced by the provider and distributed to customers. Customer intimacy, with its focus on individual customers needs, is closely aligned to this perspective. Customer intimate organizations do not propose solutions fitting most demanded market requirements, but closely cooperate with the customer in order to understand his needs and requirements, thereby providing a perfectly suited solution. Quoting executive management from a customer intimacy

driven organization, (Treacy & Wiersema, 1997, p.41) state that "the product is conceived at the customer's office". Moreover, the customer intimate organization not only provides the solution, but also ensures that, once deployed, the solution fulfills the customer expectations: they take responsibility for results. Therefore, the focus of customer intimacy is not on the value in exchange but on the value in use, as perceived by the customer.

- **The emphasis is on the process rather than on the outcome**
  The second dimension of the relationship perspective emphasizes the notion of process, consisting in multiple interactions with the customer on a long-term perspective, instead of considering the outcome of single transactions in the short-term. This view is shared by customer intimacy driven companies, for which most relevant key performance indicators are based on long-term customer lifetime value and customer retention rates rather than on market shares at a specific point in time. Indeed, customer intimacy firms invest in the customer in the initial interactions in order to grow inside the customer organization and to leverage the existing potential in its operations. Therefore, the operating model of customer intimacy driven organizations is built around the relationship process established with the customer.

- **The offering can be tangible or intangible**
  The third similarity refers to the absence of distinction based on the degree of intangibility of the value proposition in the definition of the customer intimacy. The focus is on the solution, which consists of a combination of all required elements to fulfill the customer's needs. These elements can be tangible goods as well as intangible services. In this sense, customer intimacy is close to the service-dominant logic perspective which considers that the tangibility dimension is not the most important factor: "all economies are service economies" (Vargo & Lusch, 2004a, p.10). Moreover, as in relationship marketing, customer intimacy insists on the previously presented the degree of individualization: customer intimate firms rely on their

ability to customize and individualize their offering to the customer in order to achieve a competitive advantage.

- **Knowledge is the main source of competitive advantage**
  Finally, the emphasis on knowledge is the fourth commonality between these concepts. While the service-dominant logic states that knowledge in a broad sense is the fundamental source of competitive advantage, customer related knowledge, insight in the customer business and the ability to use them are key differentiators of customer intimacy and relationship marketing. Indeed, close similarities can be found among the following two statements. Focusing on customer intimacy, Treacy & Wiersema (1997, p. 131) argue that "deep customer knowledge and breakthrough insights about the client's underlying processes are the backbone of every customer-intimate organization." Focusing on relationship marketing, Grönroos (2007, p.30) considers that a "key requirement in relationship marketing strategy is that a manufacturer, wholesale, retailer, service firm, or supplier knows the long-term processing needs and desires of their customer better and offers value on top of the technical solutions embedded in consumer goods, industrial equipment or services." In addition, Berry (1995, p.153) confirms that "relationship marketing allows service providers to become more knowledgeable about the customer's requirements and needs." It is established, thus, that customer intimacy, relationship marketing, and the service-dominant logic all recognize the significance of knowledge and specifically customer related knowledge.

In conclusion, this analysis demonstrates that customer intimacy is a type of strategy which is strongly related to services, closely aligned with relationship marketing, and which shares multiple similarities with the service-dominant logic. In the next section, this thesis describes three approaches related to the adoption of the customer intimacy value discipline.

2.3. Three Approaches Related to Customer Intimacy

This thesis aims at providing a model and a methodology for assessing and monitoring customer intimacy in B2B markets and, therefore, to support the relationship marketing activities of B2B providers. In that sense, this contribution relates to existing approaches for adopting a marketing strategy. The objective of this section is to elaborate on the similarities between customer intimacy and three marketing approaches, namely key account management, market orientation, and customer relationship management. While key account management focuses on individual relationships with the most important customers in a B2B context, market orientation defines a culture centered around the management of customers' and competitors' related knowledge, and customer relationship management allows the organization to focus on most profitable business relationships.

2.3.1. Key Account Management

The concept of key account management has emerged over the last 40 years along with the development of relationship marketing. In the literature, it was also referred to as large account management, global account management, or strategic account management (Holt & McDonald, 2000; Boles *et al.*, 1999). The rationale of key account management is to develop a specific marketing program for the provider's most important customers in the context of B2B markets. If a limited number of customers generate the most important share of revenues and profits, it is sound to allocate dedicated employees and teams to focus exclusively on the management of the relationships with these customers. According to Cannon & Narayandas (2000, p.408), key account management is the "embodiment and implementation of the relationship marketing paradigm for large business customers." Wengler (2006, p.27) defines key account management as "a supplier's relationship marketing program which aims at establishing, developing and maintaining a successful and mutually beneficial business relationship with the company's most important customers."

From the perspective of the provider, the main objectives of key account management are to ensure customer retention and to maximize the customer value (Wengler, 2006; Cannon & Narayandas, 2000; Berger *et al.*, 2002).

Customer retention means to keep the customer and to ensure that he will generate regular incomes over time, for instance by purchasing products or services every quarter or every year. Customer retention has, thus, been established as an indicator of the loyalty of the customer to the provider (Lam *et al.*, 2004). The motivation for focusing on customer retention builds upon empirical evidence which demonstrates that it is cheaper to keep a customer rather than to acquire a new one. Reichheld & Sasser (1990) established that a 5% increase of the customer retention rate can generate up to 85% improvement in profitability. In order to retain customers, two means have proven to be successful. Providers can either improve customer satisfaction or increase switching costs:[13] "both enhancing customer satisfaction and increasing switching costs can be seen as important strategies that promote customer loyalty" (Lam *et al.*, 2004, p.308). Consequently, the key account manager responsibilities can be derived from the objective of retaining customers: he should ensure customer satisfaction by providing solutions that fulfill the customer needs and expectations, as well as try to increase the switching costs by making the customer more dependent on the provider capabilities, skills and knowledge.

The second objective of key account management, maximizing customer value, is derived from various analysis establishing that customer retention is not a sufficient condition for being successful: the business relationships must be profitable (Reinartz & Kumar, 2000). The important resources committed by the provider, with specific teams focusing on individual customers, have to lead to a positive return on investment in the long run. Therefore, a significant contribution of key account management to relationship marketing literature lies in the definition and assessment of different indicators to assess

---

[13]   Switching costs are the costs incurred to the customer when changing the supplier (Lam *et al.*, 2004).

this degree of "long-term profitability" value, such as customer lifetime value, customer equity, and return on relationships. Customer lifetime value is a monetary approach of the overall value returned by the customer to the provider. In this perspective the customer is seen as any other investment of the provider, and the customer lifetime value is calculated as the net present value of the contribution margin over the relationship lifetime (Berger *et al.*, 2002).[14] Customer equity enlarges this measurement and aggregates customer lifetime value over all actual and potential customers of the provider in the industry.[15] Finally, return on relationships is estimated from a network perspective and measures the net financial outcome of the overall relationship network of the provider.[16] Consequently, in order to maximize customer value, key account managers are responsible for minimizing the costs of the relationship, for instance by reducing the process and transaction costs and by removing uncertainty to make business relationships more predictable. They are also responsible for expanding the provider's business activities inside the account, by identifying new opportunities for partnership and synergies with the customer (McDonald *et al.*, 1997).

A characteristic of marketing in B2B markets is that it involves many individuals from the provider and the customer organizations.[17] People with diverse functions, knowledge and skills on both sides participate in the relationship process. For instance, sales employees actively communicate with the purchasing department and the head of the customer organization. Services employees cooperate with various customer employees in order to perform their tasks. Therefore, multiple interactions occur within the scope of the relationship

---

[14]   This is calculated as the sum of the discounted earnings (revenues minus costs) over the lifetime of the relationship (Berger *et al.*, 2002).

[15]   Rust *et al.* (2004, p.110) define customer equity as "the total of the discounted lifetime values summed over all of the firm's current and potential customers."

[16]   Gummesson (2008, p.257) defines return on relationships as "the long-term net financial outcome caused by the establishment and maintenance of an organization's network of relationships."

[17]   "The many-headed customer and the many-headed supplier" is the 6th element out of the 30 Rs of relationship marketing (Gummesson, 2008, p.91).

and a network formed by provider and customer employees has to be coordinated. This coordination task is an essential aspect of the key account manager's activities (Holt & McDonald, 2000). Acting at the interface between both companies, the key account manager represents the customer inside the provider organization and embed the customer as far as possible in the provider's own processes. On the other side – inside the customer organization – the key account manager coordinates the provider resources, optimizes their utilization, and ensures that a clear communication is established between provider and customer employees.

This notion of interaction based relationship network is foundational for the contribution of this thesis. Chapter 5 introduces the *CI Analytics* model to infer this relationship network by applying machine learning algorithms on customer related data. This model is complementarity to key account management: the solution proposed by this thesis and prototypically implemented in the software *CI Analytics*[18] supports key account managers with regard to their investments decisions and help them coordinate this relationship network.

2.3.2. Market Orientation

The concept of market orientation has originally been proposed in order to elaborate the actual steps required to implement the marketing strategy, instead of considering marketing as a "business philosophy" (Deng & Dart, 1994).[19] Indeed, the focus of market orientation is on specifying a set of activities that a firm should perform to achieve its marketing objectives, rather than on defining the concept of marketing itself. Market orientation modifies the firm behavior with regard to its customers and competitors, and also influences its organizational structure. This notion has emerged in marketing literature as several studies proved the positive impact of adhering to

---

[18]   This software is described in chapter 6.
[19]   Deng & Dart (1994, p.726) define the marketing concept as a business philosophy holding that "long term profitability is best achieved by focusing the coordinated activities of the organization toward satisfying the needs of a particular market segment."

market orientation on business performance (Narver & Slater, 1990; Kohli & Jaworski, 1990; Rodrigez Cano *et al.*, 2004).

Even though this concept was defined in multiple ways, Jaworski & Kohli (1993, p.54) introduced a definition of market orientation which is recognized as a reference and which consists of the three following aspects: "(i) organization-wide generation of market intelligence pertaining to current and future customer needs; (ii) dissemination of the intelligence across departments; (iii) the organization-wide responsiveness to it." The first aspect - generation of market intelligence – refers to the ability of the organization to acquire three different categories of knowledge: knowledge about the customer needs and preferences, knowledge about competitors and their ability to fulfill these needs, and finally knowledge about the customer market and environment, which might influence the customer behavior, such as government regulations. The second aspect – intelligence dissemination – refers to the ability of the entire organization to share this acquired knowledge in a way that reaches the employees who can use it. In order to achieve this, the firm has to establish both vertical and horizontal communication structures so that all departments, teams, and employees can easily exchange relevant market intelligence information. The third aspect – responsiveness – refers to the action taken in response to the acquired market intelligence. Gathering and exchanging market intelligence information does not improve the created value, the competitiveness, or the business performance unless this knowledge is actually leveraged. In order to react on this knowledge, the firm can, for instance, choose to focus on specific market segments. It can also promote its offering in a way that create some interest in the customer organization, or adapt its products and services to anticipate the customer needs.

According to Narver & Slater (1990) and Deng & Dart (1994), the firm has to focus on four main dimensions in order to achieve market orientation: customer orientation, competitor orientation, interfunctional coordination, and profit emphasis. Customer orientation represents the extent to which the firm adopt behaviors demonstrating its commitment to its customers. It refers to the ability of the firm to obtain and understand its customers needs and to provide an

adequate response ensuring the satisfaction of its customers. Competitor orientation represents the firm's ability to gather information about its competitors and to act upon it. For instance, the firm can enhance its products or services with new features in order to improve the competitiveness of its value proposition or it can modify its pricing model. Inter-functional coordination relates to the ability of the different teams and departments of the firm to collaborate, share information, and coordinate their activities in response to the acquired customer and competitor intelligence. Finally, profit emphasis reflects the ability of the firm to consider profitability as a key performance indicator.

The comparison of customer intimacy and market orientation allows to establish some similarities as well as some differences between these two concepts. Market orientation is both a broader and narrower concept than customer intimacy. The main similarity consists of the importance of knowledge and, more specifically, the emphasis on customer related knowledge in both approaches. While market orientation insists on gathering market intelligence and acting upon this information accordingly, customer intimacy focuses on obtaining knowledge about the customer's needs and expectations in order to tailor and shape the offering. Tuominen *et al.* (2004) confirm this commonality as they established a strong association between customer intimacy and market orientation. Moreover, both concepts emphasize the need to involve the entire organization, and not only the marketing department in the process of managing customer related knowledge: market orientation requires a strong ability to disseminate market intelligence and well established "inter-functional coordination". Similarly, the customer intimacy operating model requires that all entities of the organization are focused on solving customers' problems and empowers the employees in contact with the customer.

However, market orientation is different from customer intimacy as it does not focus on the customer only, but on the overall market intelligence and includes also knowledge related to the firm's competitors. The objective of market orientation is not to fulfill to the highest extent the needs and expectations of individual customers, as customer intimacy does, but to understand these needs, to understand

the competitive offers available on the market, and to provide a solution which is better than those of competitors. In addition, in market orientation, the emphasis is solely on knowledge and acting upon this knowledge: it does not consider the relationship established between the customer and the provider. As opposed to customer intimacy, market orientation is not grounded in relationship marketing: the objective is not to involve the customer as a partner to co-create the value. In market orientation, the provider gathers market intelligence and act upon it in order to improve its value proposition for a specific market segments, but customers do not participate directly to the design of this value proposition on an individual basis. Moreover, even though some articles related to market orientation refers to its long-term focus, this is to outline the long-term sustainability of the firm, rather than the development of long-term relationships with customers (Narver & Slater, 1990). The focus of market orientation remains the outcomes produced by the firms for its customers rather than the processes of value creation with its customers.

This comparison of the concepts of market orientation and customer intimacy leads to the conclusion that customer intimacy cannot be assessed in the same way as market orientation is measured. Indeed, while some of these aspects related to the customer related knowledge can be taken into consideration for the evaluation of customer intimacy, the assessment of customer intimacy must include the customer relationship dimension.

2.3.3. Customer Relationship Management

The concept of customer relationship management (CRM) has become popular in the late 1990s, mainly through its association with IT, and more specifically with the development of IT based CRM systems, which aim at supporting the management of the relationships with customers and their underlying interactions. Several software providers and consulting firms have included CRM in their portfolio, and this market represents in 2010 over \$10B (Maoz *et al.*, 2010). However, CRM cannot be reduced to this technological perspective without the risk to jeopardize the CRM initiative. Indeed, the fact

that firms perceive CRM only as a technological project is seen as a significant reason for the failure of CRM adoption (Doherty & Lockett, 2004). In order to successfully achieve CRM, a change in the mindset of the organization is required. Hasan (2003, p.16) argues that to adopt CRM, "companies must make a fundamental change in the way they do business, modifying their approach to sharing information and coordinating activities within the company."

From the technological point of view of software vendors to the philosophical approach of CRM, considering it as a "business mindset", the CRM concept has been investigated in numerous ways. In a thorough literature review, Zablah *et al.* (2004) identified five dominant CRM perspectives: strategy, process, philosophy, capability, and technology. This section focuses on the strategic and operational – process based – perspectives in order to outline the commonalities of CRM with relationship marketing and customer intimacy.

From a strategic perspective, Payne & Frow (2005, p.168) define CRM as "a strategic approach that is concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments." A similarity can be perceived between this definition and the two dimensions of the previously described relationship perspective: both this definition and the relationship perspective emphasize the process of developing customer relationships as well as the creation of value for all shareholders, including both the provider and the customer. Parvatiyar & Sheth (2001) recognize that the terms CRM and relationship marketing have been often used to describe the same phenomenon. More precisely, the association between relationship marketing and CRM is in somehow similar to the association between market orientation and marketing: while market orientation is defined as the implementation of the marketing concept, CRM is described in literature as the means to adopt relationship marketing. Zablah *et al.* (2004, p.480) confirm that "relationship marketing is often cited as the philosophical basis of customer relationship management". Gummesson (2008, p.7) further insists on the practical aspects of CRM and defines it as "the values and strategies of relationship marketing [...] turned

into practical application and dependent on both human action and information technology."

From the operational perspective, much literature has focused on defining CRM as a set of processes. Reinartz *et al.* (2004, p.294) define CRM as "a systematic process to manage customer relationship initiation, maintenance, and termination across all customer contact points in order to maximize the value of the relationship portfolio." This is a broad perspective which covers the life cycle of the relationship and which is closely aligned to the definition of relationship marketing presented in section 2.2.1.2. Bueren *et al.* (2004) and Gebert *et al.* (2003) further detail this process perspective and argue that CRM consists of six sub-processes:

- *Campaign management* refers to the segmentation of the market in smaller groups of customer and prospective customers, and then, to the planning and realization of customized communications and interactions with these targeted groups of customers.

- *Lead management* refers to the systematic identification and prioritization of potential sales opportunities which raise customers' interest.

- *Offer management*, as the core sales activities, relates to the process of qualifying leads with the customer and transforming them into offers that the customer can purchase.

- *Contract management* is the process of maintaining and adjusting long-term contract in order to ensure that customers' expectations remain fulfilled, even in the case that customers' needs have changed.

- *Complaint management* ensures that all issues encountered by customers as well as all sources of dissatisfaction are actually tracked and managed consistently.

- *Service management* focuses on the maintenance, repair, and support activities related to the customers' purchases.

These descriptions of the strategic and operational perspectives on CRM outline the close association between this concept and relationship marketing. They also highlight an important characteristic of CRM which distinguishes it from customer intimacy. In contrast to customer intimacy, CRM does not focus on customizing the value proposition and adapting the offering in order to fit exactly the needs of each customer. The goal is not to establish close and collaborative relationships that tend to transform in partnership with all customers. On the contrary, the multiple CRM definitions insist on the objective to create value for the shareholders and to maximize it by targeting the most profitable customers. In that sense, CRM does not exclude transactional relationships as long as they remain profitable. Such relationships may not generate as much revenue as closer ones, but they also require a smaller investment in time and resources and, thus, may be profitable. Zablah *et al.* (2004, p.481) confirm that "CRM is concerned with the development and maintenance of a portfolio of profit-maximizing customer relationships that is likely to include exchange relationships that vary along the transactional-relational continuum." This characteristics has two main consequences: it impacts the target of the CRM initiative and lowers its emphasis on customer related knowledge:

- Since CRM allows to some extent transactional and non collaborative relationships, its target includes all customers and prospective customers of the firm. Ryals & Knox (2001, p.535) confirm that "CRM provides management with the opportunity to implement relationship marketing on a company-wide basis." While relationship marketing emphasizes the relationship and interaction with individual customers, CRM provides the firm with the ability to focus on the entire market. Plinke (1997, p.19) categorizes CRM as a relationship marketing program targeted on the market or some of its segments. The close association between IT and CRM is derived from this aspects: firms rely on technology in order to manage, support, and even individualize the interactions with customers.

- Since CRM is not focused on the individualization of the value proposition, it also has a smaller emphasis on customer re-

lated knowledge than customer intimacy. The previously proposed strategy focused definition of CRM does not mention customer knowledge and its management. In the six previously described CRM subprocesses, customer needs, as a form of knowledge about the customer, are only mentioned in the offer and the contract management. These processes, however, do not detail the management and dissemination of customer knowledge. Gibbert *et al.* (2002) argue that CRM is only focused on knowledge *about* customers: customer relationship management mines knowledge *about* customers in order to achieve customer retention, but does not consider knowledge *from* customers in order to improve the value proposition for the customer.

The next section summarizes the results of the analysis of the concept of customer intimacy performed in this chapter.

2.3.4. Customer Intimacy: A Specific Adoption of the Marketing Concept

In the previous sections, three concepts closely related to, but distinct from, customer intimacy have been introduced: key account management, market orientation, and customer relationship management. The commonalities and differences between these marketing endeavors and customer intimacy have been outlined and can be summarized along the following three dimensions, as depicted in table 2.1: primary objective of the program, focus on customer relationships, and focus on customer knowledge.

- **Primary objective of the program**
  The first dimension refers to the objective of the marketing initiative. While the primary objective of customer intimacy is to achieve a competitive advantage through the individualization of the value proposition and the fulfillment of customer needs, thereby providing the best solution to the customer, key account management focuses on retaining the most important

Table 2.1.: Comparison of Customer Intimacy With Other Marketing Programs

	Customer Intimacy	Key Account Management	Market Orientation	Customer Relationship Management
Primary objective of the program	Best solution (for all customers)	Customer retention and customer value maximization (for selected customers)	Profitability and market position improvement	Profitability of the relationship portfolio
Focus on customer relationships	++	++	−	+
Focus on customer knowledge	++	++	++	+

customers and on maximizing their value, CRM aims at achieving a portfolio of profitable relationships, and market orientation considers the overall profitability of the firm and its position on the market. Customer intimacy is more focused on the individualization of the value proposition than key account management and CRM because the entire customer intimate organization is structured around the objective to provide a solution fitting the requirements of the customer, whereas in the case of key account management and CRM, the individualization of the value proposition is achieved only if this is necessary to keep the customer and if this is profitable from a long-term perspective. With regard to market orientation, the individualization of the offering is perceived as a means to respond to the acquired market intelligence. It is only performed if it improves the overall profitability of the firm and its position

on the market.

- **Focus on customer relationships**
  The second dimension refers to the establishment of relationships with customers. Since customer intimacy, key account management, and CRM are all grounded in the concept of relationship marketing, these three concepts focus on the establishment of customer relationships. They are in particular a key requirement for the successful implementation of customer intimacy and key account management. Market orientation, however, is different and has a lower emphasis on relationships. Relationships are perceived, in the context of market orientation, as a means to acquire customer knowledge. Indeed, a market orientation program can be carried out in a transactional perspective.

- **Focus on customer knowledge**
  The third aspect is concerned with the management of customer knowledge. CRM has a lower emphasis on customer knowledge than key account management and customer intimacy, as it primarily focuses on knowledge about the customer, and more precisely on mining this knowledge. On the contrary, key account management and customer intimacy consider customer knowledge as a fundamental source of competitive advantage and develop a stronger emphasis on its management. Customer knowledge is also a central aspect of market orientation. Market orientation, however, also focuses on knowledge related to competitors in order to determine the market position of the firm.

In conclusion, customer intimacy can be perceived as a highly developed implementation of the concept of relationship marketing with a high focus on establishing customer relationships, on managing customer knowledge, and on leveraging these two aspects in order to derive competitive advantages. Moreover, its closeness to the main service dimensions and to the service-dominant logic makes it a very well suited strategy for all organizations which are going through a servitization endeavor.

3. Methods and Techniques to Assess Customer Intimacy

The objective of this chapter is to introduce the methods and techniques leveraged in this thesis in order to perform the assessment of customer intimacy in a Business to Business (B2B) context, namely network analysis and data mining.

In order to achieve the objective to provide the customer intimacy assessment along multiple degrees of granularity, from a focus on the entire customer organization to a specific analysis of customer teams and employees, this thesis proposes to apply social network analysis techniques which provides this ability to consider different entities and different levels of detail as well as to visualize the information using graph based representations. Thus, section 3.1 will introduce the concept of network analysis.

An essential part of this thesis lies in the application of data mining techniques in order to calibrate and validate the generic customer intimacy metrics presented in chapter 5. Therefore, section 3.2 will subsequently outline the main steps of the data mining process as well as the algorithms chosen in this thesis in order to perform the analysis.

3.1. Network Analysis

The application of network analysis and more specifically social network analysis in order to understand relationships among B2B organizations has already been established in past literature. Gummesson (2008, p.296) argues that network theory is "more comprehensive" in that regard than other theories such as systems or transaction costs theories because it does not focus on boundaries between the different actors, but rather on the inter-organizational aspects. Knoke & Yang (2008, p.1) confirm that the application of social network analysis in the social science literature has grown exponentially over the past 30 years, and indicate that a significant benefit of social network analysis lies in the consideration of multiple levels of analysis, defined as "individual and systemic", which allows an understanding of the "variation in structural relations and their consequences." Brandes & Erlebach (2005b) explain that three different levels of analysis are available: element-level analysis, group-level analysis, and network-level analysis. This characteristic of social network analysis allows to perform the assessment of customer intimacy at multiple levels of details, such as individuals, teams and business units, and whole organizations and, thus, confirms the relevance of social network analysis in this thesis.

Networks and more specifically social networks have been defined in numerous ways in past literature. An initial contribution to this definition is provided by Mitchell (1969, p.2) who argues that a social network is "a specific set of linkages among a defined set of persons, with the additional property that the characteristics of these linkages as a whole may be used to interpret the social behavior of the persons involved." This definition emphasizes the purpose of social network representation, which is to gain a better understanding of the actors and their relationships. More recently, Knoke & Yang (2008, p.8) presented a definition which is more focused on the inherent composition of social networks: they define a social network as "a structure composed of a set of actors, some of whose members are connected by a set of one or more relations." In the context of this thesis, the actors are the provider and customer employees and

the relations consist of the multiple relationships established through interactions and shared activities.

3.1.1. Graph Theory for the Representation of Social Networks

Graph theory has been widely adopted for the representation of social networks as the concepts of actors and relations can easily be mapped to the graph theory's notions of vertices and edges. This thesis adopts the standard graph terminology explained in Brandes & Erlebach (2005a, p.7): "a graph $G = (V, E)$ is an abstract object formed by a set V of vertices (nodes) and a set E of edges (links) that join (connect) pairs of vertices." Two vertices connected via an edge are adjacent or neighbors and are called the end vertices of the edge. It is possible to calculate the degree $d(v)$ of the vertex v by counting the number of edges in E which have the vertex v as one of their end vertices. In this thesis, the actors, which are the employees of the provider and customer organizations are represented by vertices, and the relationships established among them are represented by edges on the graph. Thus, $d(v)$ is a representation of the number of direct contacts of the employee v inside the network. Graphs can be characterized with two additional properties:

- A graph $G = (V, E)$ can be directed or undirected. If the graph is directed, the order of the end vertices of an edge is relevant for understanding the graph: the edge $e_{u,v} = \{u, v\}$ formed by the end vertices u as origin and v as destination is different from the edge $e_{v,u} = \{v, u\}$ whose origin and destination are respectively the vertices v and u. If the graph is undirected, the notions of origin and destination to qualify the end vertices of an edge become irrelevant: the vertices u and v are simply connected via the edge: $e_{u,v}$ and $e_{v,u}$ have the same meaning in the graph. In this thesis, the graphs presented in chapter 5 are undirected as the values of the calculated customer intimacy components at the individual level do not require to specify whether the end vertices of the edge are the origin or the destination vertices.

- A graph $G = (V, E)$ can be weighted or unweighted. If the graph is weighted, then a numerical value is associated to each edge on the graph. More formally, the weights can be derived by applying a weighting function $\omega : E \to \Re$. If $w_{i,j}$ is the weight of the edge $e_{i,j}$, then $w_{i,j} = \omega(e_{i,j})$. Since the same edge may have a high or a low weight depending on the chosen weighting function, this function significantly impacts the graph representation of the social networks. Moreover, using the same set of data, an infinite number of weighting functions can be derived (De Choudhury *et al.*, 2010). Therefore, the weighting function ω has to be carefully determined in order to achieve the objective of the graph representation. In this thesis, the graphs which are considered and calculated are weighted. As detailed in chapter 5, an objective of this thesis is to determine the weighting functions which provide the most accurate assessment of the values of the customer intimacy components defined in chapter 4, such as acquired customer knowledge and established customer relationships.

Two different types of matrices, the incidence matrix and the adjacency matrix provide a formal mathematical representation of the graph $G(V, E)$ (Brandes & Erlebach, 2005a). In this thesis, the adjacency matrix $A(G)$ is used by the algorithms which have been designed for the calculation of the different graphs and their customer intimacy values. The rows and columns of this matrix both represent the vertices $V = \{v_1, ..., v_n\}$ of the graph, n being the cardinality of V. Thus, $A(G)$ is a square matrix of size $n \times n$. The entry $a(i, j)$ in this matrix indicates the existence of an edge in the graph between the nodes i and j if its value is equal to 1. Otherwise, its value is equal to 0. The adjacency matrix is defined as follows: $A(G) = [a_{i,j}] \quad | \quad \forall i, j \quad 1 \leq i, j \leq n$ with:

$$a_{i,j} = \begin{cases} 1 & \text{if } e_{i,j} \in E \\ 0 & \text{otherwise} \end{cases}$$

Moreover, as described by Newman (2004), since the graphs determined in this thesis are weighted, it is also possible to calculate the

weighted adjacency matrix W.[1] In this matrix, the value of the entry $w(i,j)$ is equal to the weight of the edge $e_{i,j}$ if the edge $e_{i,j}$ exists, and to 0 otherwise. With $\omega : E \rightarrow \Re$ being the function defined to calculate the weights of the edges in the graph G, the weighted adjacency matrix $W(G)$ is defined as follows: $W(G) = [w_{i,j}] \mid \forall i,j \; 1 \leq i,j \leq n$ with:

$$w_{i,j} = \begin{cases} \omega(e_{i,j}) & \text{if } e_{i,j} \in E \\ 0 & \text{otherwise} \end{cases}$$

This thesis focuses on the representation of the customer intimacy established between two distinct entities, the provider organization P and the customer organization C, as well as between their respective employees. Thus, the graph representation of the social network investigated in this thesis has a specific topology named weighted bipartite graph. Asratian *et al.* (1998, p.7) explain that "a graph G is bipartite if the vertex set $V(G)$ can be partitioned into two sets V_1 and V_2 in such a way that no two vertices from the same set are adjacent." In this thesis, if V_P and V_C represent the sets of provider and customer employees, the edges of the graph $G(V,E)$ all have one end vertex in the set V_P and the other one in the set V_C: there is no edge between two nodes which belong to the same set V_P or V_C. Figure 3.1(a) illustrates such a bipartite graph representation with the provider and customer organizations consisting of four and three employees: $V_P = \{P_1, P_2, P_3, P_4\}$ and $V_C = \{C_1, C_2, C_3\}$. The adjacency matrix of bipartite graphs has a special characteristic. As explained by Asratian *et al.* (1998, p.16), "let G be a graph with vertices $v_1, v_2, ..., v_n$ and adjacency matrix $A(G) = [a_{i,j}]$. Then G is bipartite if and only if there is a permutation Π of the set $\{1, 2, ..., n\}$ so that the matrix $A'(G) = [a_{\Pi(i),\Pi(j)}]$ has the following form:

$$\begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$$

---

[1] the adjacency matrix is sometimes called binary adjacency matrix to differentiate it from the weighted adjacency matrix (Kiss, 2007, p.72).

where B^T is the transpose of B." Indeed, as depicted in figure 3.1(b), the adjacency matrix of the graph proposed in figure 3.1(a) presents such a structure.

	P1	P2	P3	P4	C1	C2	C3
P1	0	0	0	0	2	0	0
P2	0	0	0	0	1	0.7	0.5
P3	0	0	0	0	0.5	0	0
P4	0	0	0	0	0	1	2
C1	2	1	0.5	0	0	0	0
C2	0	0.7	0	1	0	0	0
C3	0	0.5	0	2	0	0	0
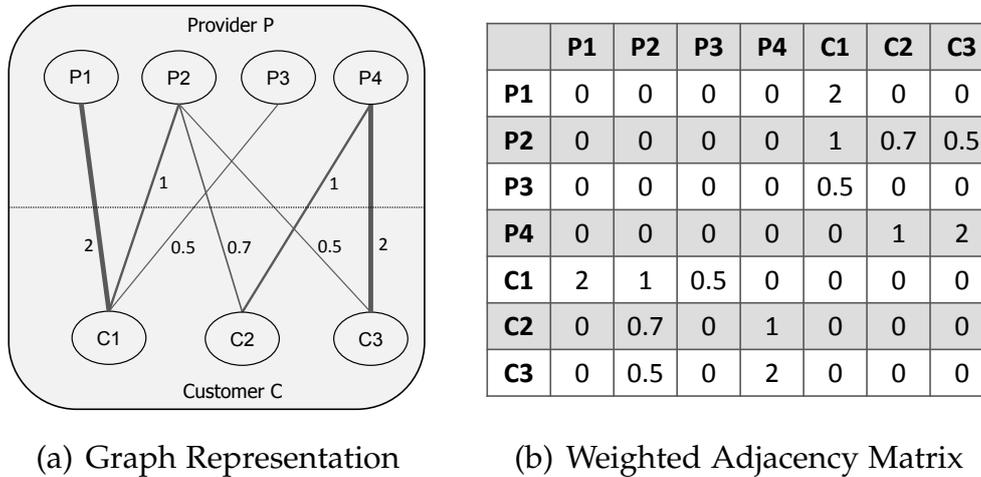
(a) Graph Representation   (b) Weighted Adjacency Matrix

Figure 3.1.: A Weighted Bipartite Graph Representation of the Provider-Customer Relationship

3.1.2. Centrality Metrics for the Analysis of Social Networks

In order to perform an analysis of the social network based on the graph representation presented in the previous section, various centrality metrics have been proposed in past literature (Freeman, 1979). Centrality metrics are particularly important for the analysis of networks as they enable an aggregation of the information presented in the graph and they provide an understanding of the relative position and importance of each actor inside the network. Many centrality metrics can be calculated in order to assess diverse characteristics of a node in the graph (Koschützki *et al.*, 2005). The following three centrality metrics have been considered in this thesis as they are well established for understanding the role and importance of each actor in the social network (Buechel & Buskens, 2008; Freeman, 1979):

1. **Degree Centrality**
   Degree centrality is one of the first centrality metrics which has

been conceived and is, in its first definition, a synonym of the previously defined notion of degree (Koschützki *et al.*, 2005). The degree centrality $C_D(i)$ of the vertex i in the graph $G(V, E)$ indicates the number of adjacent vertices to i, or the number of edges which have i as one of their end vertex. Considering the previously defined adjacency matrix $A(G)$ of the graph G and n being the cardinality of V, $C_D(i)$ is calculated as follows:

$$C_D(i) = \sum_{j=1}^{n} a_{i,j} \qquad (3.1)$$

In order to make the degree centrality comparable among graphs of different sizes, a normalized form of the degree centrality has been proposed, in which the degree centrality is divided by the maximum number of potential neighbors on the graph (Freeman, 1979; Wasserman & Faust, 1994). With n being the cardinality of V in the graph $G(V, E)$, the normalized degree centrality $C'_D(V)$ is calculated as follows:

$$C'_D(i) = \frac{\sum_{j=1}^{n} a_{i,j}}{n-1} \qquad (3.2)$$

The degree centrality and normalized degree centrality are indications of the neighborhood of the actors in the network as they specify the numbers of actors which can be directly reached. In this thesis, since the calculated graphs are bipartite, these two centrality metrics indicate the numbers of relationships established by a provider (resp. customer) employee inside the customer (resp. provider) organization.

2. **Closeness Centrality**
   The closeness centrality $C_C(i)$ reflects to which extent the vertex i is near or far from the other nodes on the graph. Sabidussi (1966) proposed a first calculation of the closeness centrality based on the notion of distance $d_{i,j}$ between two vertices i and j. This distance $d_{i,j}$ is calculated as sum of weights of the edges

that belong to the so called geodesic or shortest path that connect i and j. Using this measure, the closeness centrality is defined as follows:

$$C_C(i) = \frac{1}{\sum_{j=1}^{n} d_{i,j}} \qquad (3.3)$$

As for the degree centrality, a normalized version has been proposed in order to remove the variation due to network size effects (Freeman, 1979; Wasserman & Faust, 1994):

$$C'_C(i) = \frac{n-1}{\sum_{j=1}^{n} d_{i,j}} \qquad (3.4)$$

In this thesis, since the customer intimacy graphs are weighted and bipartite, the geodesic distance between i and j is simply equal to the weight $w_{i,j}$ of the edge $e_{i,j}$. Thus, the closeness and normalized closeness centrality metrics are calculated as follows:

$$C_C(i) = \frac{1}{\sum_{j=1}^{n} w_{i,j}} \qquad (3.5)$$

$$C'_C(i) = \frac{n-1}{\sum_{j=1}^{n} w_{i,j}} \qquad (3.6)$$

3. **Betweenness Centrality**
   The third important centrality metric is called betweenness centrality. Its objective is to indicate the relative importance and power of control of each vertex of the graph. Vertices that have a high betweenness centrality are located on a high number of geodesic paths that connect the other nodes in the graph. Since the graphs considered in this thesis are bipartite, and because the focus is only the relationships between provider and customer employees, there is no vertex on the graph which is located on other vertices' geodesic path. As a consequence, this metric is not relevant in this thesis and, thus, not further detailed in this section. Additional information on this metric can be found in Koschützki *et al.* (2005, p.29).

3.1.3. Using Social Network Analysis for Assessing Customer Intimacy

In the previous sections, the notion of a social network, its representation in the form of a graph containing vertices and edges, as well as its analysis by means of centrality metrics have been explained. In order to perform an analysis of the social network, it is also necessary to explicit the meaning of the relational ties that exist between actors in the network, and which are represented by edges between the vertices of the graph. Wasserman & Faust (1994, p.18) explain that relationship ties can indicate an extensive number of meanings such as formal associations, affiliations, behavioral interactions, or evaluations of persons by others. An original aspect of this thesis lies in the consideration of two different types of relationship ties and in their association by means of data mining techniques in order to calibrate the model to assess customer intimacy. The two types of relationship tie considered in this thesis are the following:

- **Behavioral interaction**
  When the relationship ties indicate some behavioral interaction, the weight of each tie is derived from past communications and activities that occurred between the two actors related to the tie. In that case, the data collected to design the social network consists of past observations or archival records (Wasserman & Faust, 1994, p.49). Following this approach, it is explained in chapter 5 how data contained in the provider's information system is collected in order to calculate multiple customer intimacy metrics based on behavioral interaction.

- **Evaluation of one person by another**
  When the relationship ties indicate some evaluations of persons by others, the actors in the social network are asked to answer a set of questions related to other actors. These questions should reflect the objective of the social network representation which is, in this thesis, the assessment of various customer intimacy components. The data is collected either by means of interviews or through the completion of a questionnaire by the respondents. For scalability reasons, the questionnaire option has

been chosen and a "customer intimacy questionnaire" has been conceived in the course of this thesis. While the actual content of the questionnaire is introduced in chapter 5, the design characteristics of this questionnaire are outlined in the following paragraphs.

The questions asked to the respondents can either reflect a "complete ranking" or a "rating" of the relationship ties (Wasserman & Faust, 1994, p.47). In the complete ranking approach, the respondents are asked to order or to prioritize the different ties on the network with regard to a specific attribute. For instance, the respondents are asked to rank the relationships they have established with different customer employees. In the rating approach, the different relationship ties are considered independently from each other and the respondents are asked to assess the different ties on a certain scale. For instance, the respondents are asked if their relationships with different customer employees are low, medium, or high. Since "ranking" the different customers and their employees is out of the scope of this thesis, the rating approach has been chosen in order to assess the customer intimacy components.

In order to design this "rating" customer intimacy questionnaire, the well established approach based on Likert-type scales has been followed. Miller & Salkind (2002, p.330) explain that a Likert-type scale is a "summated scale consisting of a series of items to which the subject responds." These items are presented in the form of assertions for which the respondent evaluates the intensity of his agreement or disagreement by selecting a value comprised between one and seven.[2] In order to ensure the validity of the Likert-type scales developed in this thesis, the different series of items created to assess the customer intimacy components have been conceived upon past literature and previously created questionnaires which are mainly rooted in the field of relationship marketing. These are further detailed in chapter 5.

---

[2]   Some Likert-type scales are based on a different number of intensity grades such as five, six, or ten.

An important characteristic of Likert-type scales for the rest of this thesis is the nature of the scale itself. There is indeed some discussion about whether Likert-type scales should be considered as ordinal or interval scales.[3] As explained by Jamieson (2004), Likert-type scales are in their essence ordinal, even though several researchers use them as interval scales. Thus, within the scope of this thesis, the designed Likert-type scales are considered as ordinal scales. As described in section 3.2, this characteristic influences the selection of data-mining algorithms used for calibrating the model.

Further information about social networks, and more specifically about their actual application in this thesis is provided in chapter 5. The next section introduces the data-mining approach used in this thesis for calibrating the assessment of the customer intimacy components.

3.2. Data Mining

Since the calibration of the customer intimacy assessment presented in chapter 5 and applied in chapter 7 is based on data mining techniques and methods, the objective of this section is to introduce the underlying data mining concepts which are relevant for this thesis.

Part 3.2.1 introduces the process of "Knowledge Discovery in Databases" (KDD) proposed by Fayyad *et al.* (1996b), and on which the *CI Analytics* methodology elaborated in section 5.1 is aligned. This part subsequently elaborates on the concepts of data-mining and machine learning and puts them in relation to the KDD process. Part 3.2.2 motivates the selection of machine learning algorithms considered in this thesis and shortly describes them. Finally, part 3.2.3 details the means used for validating data-mining models and, thus, for confirming the overall approach proposed by this thesis to assess customer intimacy.

---

[3] An explanation of the difference between ordinal and interval scales is proposed in Hair *et al.* (2010, p.5).

3.2.1. The Process of Knowledge Discovery in Databases

With the exponential increase of data created, stored, and used over the past decades, in part due to the rise of internet and new information and communication technologies, new solutions have been required in order to analyze data and to extrapolate some sense out of it. Thus, the development of solutions, methods, and techniques for transforming data into actionable and more compact forms of information and knowledge has received considerable interest in both academia and practice. This overall process of leveraging this data to generate some knowledge has been called the Knowledge Discovery in Database (KDD) process. Fayyad *et al.* (1996a, p.6), who originally introduced this notion, define it as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data." Fayyad *et al.* (1996b, p.37) argue that this process enables "mapping low-level data into other forms that might be more compact, more abstract, or more useful." This process comprises of six different steps which are depicted in figure 3.2:

1. **Problem Definition**
   The first step in this process consists of obtaining a thorough understanding of the investigated problem and its context, as well as in identifying the sources of data which are relevant for providing a solution. As explained in the introduction, the objective of this thesis is to find some patterns in customer related data available in the provider's information system in order to perform an assessment of the degree of customer intimacy established with different customers.

2. **Selection**
   The second step relates to the selection of the data records on which the analysis will be completed and to the identification of the actual fields in the data set that will be considered.

3. **Pre-Processing**
   The third step concerns cleaning the data, such as removing noise and outliers which may prevent from identifying the patterns, and handling the missing values in the data set.

4. **Transformation**
   In the fourth step, the data is transformed in order to empha-
   size its most important characteristics. This involves the aggre-
   gation of the variables in the data set to create summated scales
   as well as the projection of the data in orthogonal dimensions
   in order to reduce the number of variables.

5. **Data Mining**
   The fifth step refers to the analysis of the data itself through the
   application of various machine learning algorithms. This step
   is further detailed in the next paragraph.

6. **Interpretation/Evaluation**
   Finally, the last step consists of the validation of the model in
   order to ensure that it can be used with other data sets as well as
   in its interpretation in order to derive some theoretical or prac-
   tical knowledge. This step is further detailed in section 3.2.3.
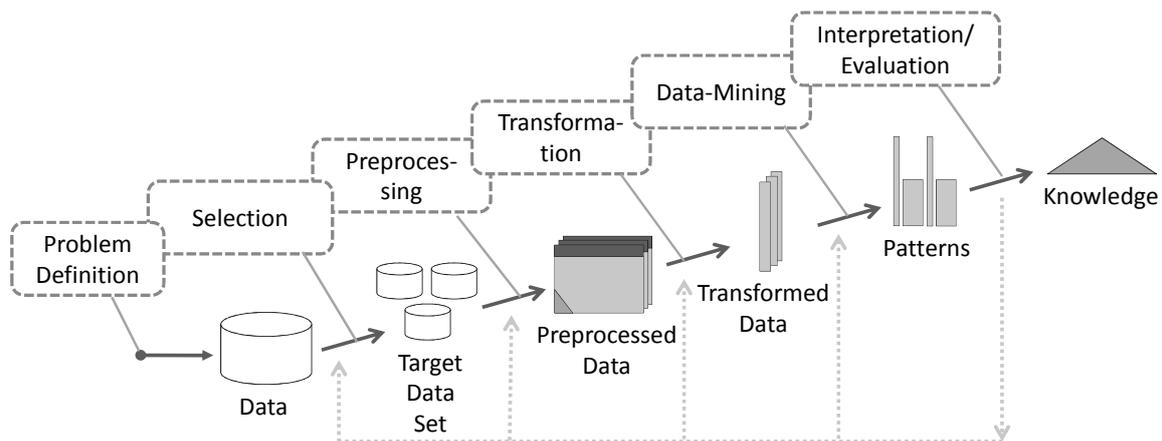
Figure 3.2.: The Knowledge Discovery Process (Fayyad *et al.*, 1996b)

Within the KDD process, the fifth activity is concerned with the ana-
lysis of the data itself and more precisely with the detection of pat-
terns among the multiple data records. This aspect is referred to
as data-mining (Witten *et al.*, 2011). Fayyad *et al.* (1996b, p.41) con-
firm that "data mining is a step in the KDD process that consists of
applying data analysis and discovery algorithms that produce a par-
ticular enumeration of patterns (or models) over the data." Notably,

since data-mining is certainly the most significant activity of the KDD process, the concepts "knowledge discovery in database" and "data mining" are sometimes used as synonyms (Kiss, 2007, p.12, Mitchell, 1999). For instance, the Cross Industry Standard Process for Data Mining (CRISP-DM) consists of 6 steps[4] which are to a high extent aligned to the KDD process (Chapman *et al.*, 2000).

Several computer-science based methods and algorithms have been conceived in order to perform the analysis of the data. These are called machine learning algorithms. As explained by Alpaydin (2010, p.3), "machine learning is programming computers to optimize a performance criterion using example data or past experience [...] Their application to large databases is called data mining." Such algorithms are rooted in the field of artificial intelligence as they have to be able to adapt to changing environments. The principle of machine learning is that the algorithm is applied to a set of data records called training set in order to create a model consisting of multiple patterns which present a structural description of the data set (Witten *et al.*, 2011, p.8). After being validated, this model can be applied on other data sets in order gain new insight. There are two different types of machine learning algorithms:

- **Unsupervised Learning**
  In the case of unsupervised learning, no specific field in the data is considered as a reference: all fields are input data and the objective is simply to identify regularities in the input (Alpaydin, 2010, p.11).

- **Supervised Learning**
  In the case of supervised learning, an attribute of the dataset is considered as the target or as the output variable of the algorithm. The algorithm is applied on the training set in order to "learn" the value of this attribute based on all other fields, which are called the input variables of the algorithm: "the task is to learn the mapping from the input to the output" (Alpaydin, 2010, p.9). Classic types of supervised learning are regres-

---

[4] These steps are (1) Business Understanding, (2) Data Understanding, (3) Data Preparation, (4) Modeling, (5) Evaluation, (6) Deployment.

sion and classification. If the target variable on the dataset is numeric and continuous, then a regression is performed: the supervised algorithm aims at creating a model which predict as closely as possible the value of the target field, based on the available input fields. If the target variable is nominal or ordinal, then a classification is performed: the algorithm aims at creating a model that predict the class or the order of the record specified in the target field, based on the other input fields.

In this thesis, the objective is to use the customer related data available in the provider's information system in order to predict the customer intimacy component values which have been empirically assessed. Consequently, the supervised machine learning approach is followed as the target variable is derived from the empirical results. As explained in section 3.1.3, this empirical analysis of the customer intimacy components is performed using ordinal Likert-type scales. Thus, from a machine learning perspective, the aim of this thesis is to perform a classification.

3.2.2. Selection of the Machine Learning Algorithms

Four classification algorithms have been considered in this thesis. While the first part of this section motivates their selection, the remaining parts briefly describe them.

3.2.2.1. Choosing Relevant Algorithms

Multiple machine learning algorithms are available in order to solve a classification problem. Many of them use, to different degrees, concepts rooted in classic inferential statistics and in Bayesian decision theory.[5] Indeed, Witten *et al.* (2011, p.28) confirm that there is no strict difference between machine-learning and statistics but "a continuum of data analysis techniques." However, in contrast to classic inferential statistics which require the dataset to fulfill certain

---

[5]   The Bayesian decision theory focuses on the estimation of class probabilities, knowing certain conditions apply or certain observations were
made (Alpaydin, 2010, p.3, p.48).

conditions, such as normality, homoscedasticity, and linearity, many machine learning algorithms have been designed for data contained in databases, which in most cases violate these conditions (Hair *et al.*, 2010, p.72, Press, 2003, p.6). The algorithms which can be applied to data whose distribution is unknown are called non-parametric algorithms (Alpaydin, 2010, p.164). Since it cannot be assumed that the customer related data available in the provider's information system follows a specific distribution, only non-parametric machine learning algorithms are considered in this thesis.

Several factors influence the performance of a machine learning algorithm on a specific data set such as the number of target classes, the distribution of the target class, the total number of cases and attributes, and the average number per class (Nisbet *et al.*, 2009, p.257). Moreover, there is no absolute analytical rule for determining the most relevant algorithms upon certain characteristics of the dataset (Kalousis *et al.*, 2004, Kiss, 2007, p.23). Thus, several projects were conducted over the past decades in order to empirically assess the performance of data-mining algorithms. In this thesis, the machine learning algorithm selection was performed on the basis of the results from three different analyses:

- The STATLOG project is considered as one of the most exhaustive evaluation of data mining algorithms as it compares the performance of 20 classification methods on 20 different datasets (Michie *et al.*, 1994). Some of its conclusions are as follows: (i) the nearest neighbor algorithm performed very well on all datasets, even though it was the slowest on large datasets; (ii) the neural network with back-propagation algorithm obtained the highest or near highest predictive performance in nearly all cases; (iii) all decision trees had a fairly constant "average" performance across all datasets.

- Lam *et al.* (2002) benchmarked on 50 data sets their custom-made algorithm "ICPL" with the algorithms *k*-nearest neighbor, C4.5 decision tree and support vector machine. The *k*-nearest neighbor algorithm achieved the highest classification accuracy

followed by the support vector machine and the ICPL algorithms.

- Ali & Smith (2006) analyzed the performance of eight algorithms (classifiers) on 100 different datasets. No algorithm could be identified whose performance was constantly above average for all 100 datasets. The support vector machine algorithm obtained the best accuracy. The decision tree C4.5 and the neural network algorithms also obtained very good results in terms on percentage of correctly classified instances.[6]

Based on this analysis, it appears that the following algorithms are highly relevant classifiers:

1. Decision tree C4.5

2. *k*-nearest neighbor

3. Neural network with back-propagation

4. Support vector machine

Thus, these four algorithms have been considered in the scope of this thesis. The next parts of this section present further details on each of them.

3.2.2.2. Decision Tree C4.5

The machine learning algorithm C4.5 belongs to the family of decision tree classifiers, which have the advantage of being graphically representable and, thus, easily interpretable. A decision tree is "a hierarchical data structure implementing the divide-and-conquer strategy" (Alpaydin, 2010, p.187). Considering a certain data record in the database with multiple attributes, the decision tree models the classification task in multiple sequences of tests on the attributes in order to determine or predict the class of the record. The different

---

[6] The performance indicators accuracy and percentage of correctly classified instances are developed in section 3.2.3.2.

test sequences which lead to the classification are hierarchically represented in the form of a tree,[7] which consists of one root node, internal nodes, branches, and terminal leaves. In a decision tree, the nodes – also called test nodes – represent the attributes on which the tests are applied, the branches represent the different test predicates, and the terminal leaves constitute the possible classes. Dunham (2002, p.93) proposes the following formalization of a decision tree: "Given a database $D = \{t_1, ..., t_n\}$ where $t_i = \{t_{i1}, t_{ih}\}$ and the database schema contains the following attributes $A = \{A_1, A_2, ..., A_h\}$. Also given a set of classes $C = \{C_1, C_2, ..., C_m\}$. A decision tree or classification tree is a tree associated with D that has the following properties: (i) each internal node is labeled with an attribute A_i; (ii) each arc is labeled with a predicate that can be applied to the attribute associated with the parent; (iii) each leaf node is labeled with a class C_j."

The objective of decision tree classifiers, such as the C4.5 algorithm, is to induce the decision tree, which means to determine the best way of splitting the data and, thus, to identify effective and accurate sequences of tests on the attributes in order to assess the class of the different records (Tan *et al.*, 2006, p.151). C4.5 was proposed by Quinlan (1986) as a successor of the ID3 algorithm. It uses the information gain Δ_{info} in order to infer the decision tree. This information gain represents the potential increase in the information value of the decision tree that would result from extending it with an additional sub-tree. This sub-tree indicates that an additional test is required in order to lead to the classification decision. More formally, the information value is called entropy I_{info} and it measures the degree of purity of the different nodes in the tree (Tan *et al.*, 2006, p.158).[8] If m represents the number of classes, t a node in the tree, and $p(i|t)$ the fraction of records belonging to a class i at the given node t, then the

---

[7]   A tree is a special type of graph that fulfills the following two conditions: it is connected and it is acyclic (Wasserman & Faust, 1994, p.119).

[8]   Other impurity measures include Gini and classification error.

entropy $I_{info}(t)$ is defined as follows:

$$I_{info}(t) = -\sum_{i=1}^{m} p(i|t)log_2(p(i|t)) \tag{3.7}$$

If $T_{children} = \{t_1, t_2, ..., t_k\}$ represents the set of children nodes of the node t, and $N(t_i)$ the number of records associated to the node t_i, then the information gain Δ_{info} is calculated as follows:

$$\Delta_{info} = I_{info}(t) - \sum_{j=1}^{k} \frac{N(t_j)}{N(t)} \times I_{info}(t_j) \tag{3.8}$$

In order to infer the decision tree, the algorithm C4.5 creates the different nodes of the tree in an iterative manner, starting with the root node. To create the node t_i, the algorithm evaluates the potential information gain Δ_{info} obtained with each input attributes. The attribute with the highest gain is set to the test node t_i. This operation is then reapplied in order to determine the children nodes of t_i and so on, until a stop criterion such as the maximum tree depth or the minimum number of items per class is reached (Tan *et al.*, 2006, p.164).

3.2.2.3. *k*-Nearest Neighbor

The *k*-nearest neighbor algorithm belongs to the so called "lazy learners" or "instance-based learning classifiers" as it does not create an explicit model representation of the knowledge provided in the training data set (Tan *et al.*, 2006, p.223, p.226). Instead, the different records contained in the training data set are all memorized by the algorithm. When a new record r has to be classified, the algorithm calculates its distance to all records in the training set, the shortest distance indicating the highest degree of similarity. The class of r is then determined upon the classes of its *k*-nearest neighbors, for instance using a majority vote scheme.

More formally, considering a training data set $D = \{d_1, d_2, ..., d_n\}$ of size n whose instances have the set of attributes $A = \{x_1, x_2, ..., x_h\}$

of size h, each record d_i can be represented by a point in the h-dimensional space R_h. In order to assess the class of the new record r, r is also represented as a point in the space R_h and its Euclidean distance to all items in D is calculated. Then, the list D_r of the k-nearest neighbors of r is computed and their respective classes is reviewed. If the k-nearest neighbors belong to the same class C_1, then r is also set to C_1. If the k-nearest neighbors belong to different classes, then majority vote or distance-weighted voting schemes are applied in order to assess the class of r.[9]

Even though this algorithm has proven its effectiveness, a key challenge in the k-nearest neighbor algorithm resides in the appropriate determination of the number k. If k is chosen too small, there is a risk of misclassifying a record because of its closeness to one specific noisy item in the training set.[10] If k is chosen too large, then some items in the training set which are far from r and of different class may remain influential in the classification of r if they belong to the k-nearest neighbors of r.

3.2.2.4. Support Vector Machine

The support vector machine classifier belongs to the kernel machine learning algorithms (Alpaydin, 2010, p.309). It can be seen as an evolution of statistical learning theory which includes concepts derived from instance based learning (Witten *et al.*, 2011, p.192). Its strength lies in its ability to handle high-dimensional data and to consider both linearly and non-linearly separable data (Tan *et al.*, 2006, p.256). This algorithm discriminates training records pertaining to two different classes by using a subset of the training data set which are called the support vectors.

Considering a training data set $D = \{d_1, d_2, ..., d_n\}$ of instances which have the set of attributes $A = \{x_1, x_2, ..., x_h\}$ and a set of two classes $C = \{C_1, C_2\}$, these instances can be represented as points in the h-dimensional space R_h. In order to identify the support vectors, the

---

[9]   The majority vote and the distance-weighted voting are explained in Tan *et al.* (2006, p.226).
[10]   Such a problem is called over-fitting.

support vector machine algorithm determines the maximum margin hyperplane.[11] If the data is linearly separable, an infinite number of hyperplanes can be identified in R_h which discriminate the items in D of class C_1 from those of class C_2 by varying the coefficients of the equation that determines the hyperplane. The support vector machine aims at identifying the hyperplane whose distances to the nearest items of class C_1 and of class C_2 are maximized. As the sum of these two distances is called the margin of the hyperplane, the objective of the support vector machine algorithm is to determine the maximum margin hyperplane. Indeed, it has been established that "decision boundaries with large margins tend to have better generalization errors than those with small margins." (Tan *et al.*, 2006, p.257). Thus, the maximum margin hyperplane should be a better discriminant of both classes than any other hyperplane.

In the case that the data is not linearly separable, and therefore no hyperplane can be found in R_h to discriminate the items of class C_1 from those of class C_2, it is possible to perform a non-linear transformation of the space R_h and then identify the maximum margin hyperplane in this newly created space. This operation is, however, resource intensive and the transformation function is unknown (Tan *et al.*, 2006, p.272). The use of kernel functions, which "replace the transformation functions" provides the ability to search for the maximum margin hyperplane in a non-linear model directly into the original space R_n. Further details on the kernel functions are provided in Alpaydin (2010, p.320).

3.2.2.5. Artificial Neural Network – Multilayer Perceptron with Back-Propagation

Artificial neural networks are parallel information processing systems which aim at reproducing the mechanisms of biological neural

---

[11] Ostaszewski (1990, p.123) defines an hyperplane as an affine set of dimension $n - 1$ in the n-dimensional space $\Re_n$ which divides $\Re_n$ into two half-spaces. Given a set of items S in $\Re_n$ and a a boundary point (support vector) in S, the hyperplane H supports the set S in $\Re_n$ at the point (support vector) a if: (i) the point a belongs to H and (ii) S is entirely contained in one of the two half-spaces formed by H(Ostaszewski, 1990, p.129).

networks. Similarly to real neural systems in which neurons are connected to each others via axons and synapses, artificial neural networks are composed of multiples neurons or nodes which are interconnected with weighted and directed links (Tan *et al.*, 2006, p.248). In this thesis, the multilayer perceptron neural network algorithm is considered. The simpler single layer perceptron consists of a layer of input neurons which represent the attributes assessed in the classification task, one output neuron whose role is to predict the class, and directed weighted edges that connect the input neurons to the output neuron. In order to classify a record r characterized by the set of attributes $A = \{x_1, x_2, ..., x_h\}$, the input neurons of the perceptron transfer concurrently r's attribute values to the output node via the corresponding weighted edges. The output node computes the value of the perceptron y as the weighted sum of the inputs. Then, it uses an activation function s to transform y into a boolean value which can be associated to a specific class. The activation function s can be linear, sigmoid (logistic), or based on a threshold value. For instance, if $y \geq 0$ then $s(y) = 1$ and the record r is classified in C_1. If $y < 0$, $s(y) = 0$ and r is classified in C_2. The learning algorithm of the perceptron consists in feeding the network iteratively with the items of the training data set and in adjusting the weights of its edges until the classes predicted by the output node correspond to the actual classes of the training items.

Since the perceptron only has one single layer and its output neuron estimates the class based on a weighed sum of the input attributes, it uses only linear discriminants in order to perform the classification task. In order to remedy this limitation, the multilayer perceptron contains additional intermediate or "hidden" layers of neurons between the input neurons and the output neurons which provide the ability to use non-linear discriminants (Alpaydin, 2010, p.246). Figure 3.3 illustrates such a network in which the classification is performed upon three input attributes x_1, x_2 and x_3. In this example the multilayer perceptron contains one hidden layer composed of two neurons. In order to classify the record r, its attribute values are presented to the respective input neurons I_1, I_2 and I_3. The hidden neurons H_1 and H_2 compute the weighted sums of values delivered

by the input neurons using the weights $w_{i,j}$ indicated on the graph and transform them with their respective activation function s_0 and s_1. The outputs of the hidden neurons are then passed through to the output neuron O. This neuron repeats the operation of calculating the weighted sum with the appropriate weights and of transforming the value with its own activation function s_0. This value is finally used in order to assess the class of the record r.
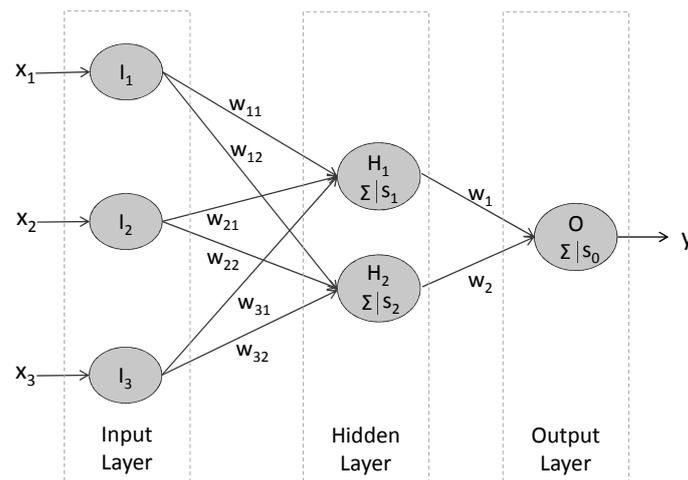
Figure 3.3.: Illustrative Multilayer Perceptron

The multilayer perceptron is a feed-forward neural network with back-propagation of the error estimate. The feed-forward characteristic indicates that the neural network is unidirectional. Indeed, as illustrated in figure 3.3, the neurons are only connected to neurons in subsequent layers which are closer to the output node (Tan *et al.*, 2006, p.251). The back-propagation of the error estimate feature indicates that the training algorithm of the multilayer perceptron is composed of multiple iterations of the following two phases: during the forward phase, the training sample records whose class are known are passed through the network iteratively in order to estimate the weights of the edges. During the backward phase, the error estimated on the sample records are transferred back to the neurons in the previous layers in order to adjust the weights. These two phases are repeated until an acceptable error estimate is reached (Tan *et al.*, 2006, p.254).

3.2.3. Evaluation of the Machine Learning Models

Once a machine learning algorithm has been trained to resolve the classification task, a machine learning model is created. This model has to be evaluated in order to ensure its ability to classify records that do not belong to the training data set and, thus, to determine its generalization error (Tan *et al.*, 2006, p.186). Different methods have been conceived in order to perform this evaluation such as holdout technique, the bootstrap, and the cross-validation. The first part of this section introduces these different options and motivates the choice to use the cross-validation technique in this thesis. In addition, several criterion have been defined in order to quantify the evaluation, like the precision and recall values or the kappa statistic. The second part of this section develops the indicators which are used to evaluate the machine learning models presented in chapter 7.

3.2.3.1. Different Options to Split the Data Set

In order to assess the capability of a machine learning algorithm to perform a certain classification task, four main techniques have been proposed (Tan *et al.*, 2006, p.186):

- **Holdout Method**
  The holdout method simply consists of splitting the data set in two subsets: a training set and a test set. Once the learning process has been performed on the training set, the resulting model is applied on the test set. The results achieved by the model on the test set are used to assess the capability of the model to determine the actual classes of records that do not belong to the training set (Tan *et al.*, 2006, p.149).[12]

- **Random Subsampling**
  The random subsampling method consists of repeating the holdout method several times: If k subsamples are created, the data set is randomly split k times, resulting in k couples of training

---

[12]   In some cases the original data set is split in a training set to create the model, a validation set to optimize it, and a test set to assess its performance (Witten *et al.*, 2011, p.149).

and test sets. The machine learning algorithm is, thus, trained and assessed k times. The overall performance of the algorithm is calculated as the average performance of the k generated models on the test sets.

- **Bootstrap**
  Similarly to random subsampling, the bootstrap method generates multiple samples from the original data set, and train and test the machine learning on each of these samples. Its specificity is that it uses a subsampling with replacement technique in order to create the sampled training data sets: any record in the original data set can be selected multiple times to compose the training set of each sample. The corresponding test set is then formed by the remaining records of the original data set which do not belong to the training set.

- *k*-**Fold Cross-Validation**
  The cross-validation technique is an evolution of the random sampling method which ensures that all records of the original data set are allocated the same number of times to the training sets and exactly once to the test sets: if k samples consisting of a training set and a test set are generated out of the original data set D, all records in D are allocated $k - 1$ times to the training sets and once to the test sets. This constraint ensures that all potential patterns in the original data set are represented in both the training and test sets. On the contrary to the bootstrap method, cross-validation does not use subsampling with replacement.

The cross-validation method has been chosen in this thesis as it is recognized as "the standard way for measuring the error rate of a learning scheme on a particular data set" (Witten *et al.*, 2011, p.154). Indeed, the other techniques all present some drawbacks: the hold-out method requires a large amount of data in order to ensure that both the training set and the test set contain sufficient representative samples. Moreover, the repartition of the data in both sets has to be

performed thoroughly since it may influence the evaluation results.[13] With the random subsampling method, the bias induced from an incorrect repartition of the data in the training and test sets is removed, but there is no control on how often records are allocated to the training and test sets, leading to a misinterpretation of the identified patterns. Finally, the bootstrapping method is particularly efficient on data set of small size. However, Witten *et al.* (2011, p.156) argues that the estimation of the error using this method is, in many cases, overly optimistic. Kohavi (1995) compared the bootstrap and the cross-validation methods on six different data sets and concluded with the recommendation to use the "10-fold cross-validation", in which the parameter k is set to the value 10.

In order to implement the k-fold cross-validation, the original data set D is partitioned in k mutually exclusive subsets of equal size and, thus, k samples are generated. The ensemble of generated subsets is defined as $R = \{R_1, R_2, ..., R_k\}$ and the ensemble of generated samples is denoted as $S = \{S_1, S_2, ..., S_k\}$. Each sample consists of test set and a training set. The test set $test_i$ and the training set $train_i$ of the sample S_i are composed respectively of the records of the part R_i and of all records which are not in R_i:

$$S_i = \{test_i, train_i\} \text{ with } test_i = R_i \text{ and } train_i = D - R_i \quad (3.9)$$

In order to evaluate the performance of the machine learning algorithm, the algorithm is trained and assessed on each of the k samples. The results achieved by the trained models on the k test sets are then combined in order to determine the overall accuracy of the algorithm.

In this thesis, the parameter k has been set to 10, as recommended in past literature (Alpaydin, 2010, p.487, Witten *et al.*, 2011, p.153, Kohavi, 1995). Thus, the original data set has been partitioned in 10 different parts and 10 samples have been generated. Moreover, as recommended by Witten *et al.* (2011, p.154), in order to ensure

---

[13]   The machine learning algorithm may have a poor or a high performance depending on whether the patterns identified in the training set also exist in the test set or not.

that the data partitioning does not bias the evaluation of the performance of the different machine learning algorithms, the entire *k*-fold cross-validation process has been repeated 10 times, each time with a different partitioning of the original data set: a "10 times 10-fold cross-validation" has been performed in order to evaluate the performance of each configuration of the different machine learning algorithms.

3.2.3.2. Model Evaluation Criteria

Several indicators have been conceived in order to assess the ability of a machine learning algorithm to solve a classification problem on a certain data set. Many of these indicators are derived from the confusion matrix. Considering a two-class classification task with the potential classes C_1 and C_2, this matrix is a 2×2 matrix as depicted in figure 3.4. The results achieved by the trained model on the records in the test set are sorted in four categories. If the class of the record is predicted as C_1 and is actually C_1, this record is classified as a true positive (tp). If the class of the record is predicted as C_2 and is actually C_2, this record is classified as true negative (tn). If the class of the record is predicted as C_1, but the record belongs in fact to the class C_2, the record is classified as false positive (fp). Finally, if the class of the record is predicted as C_2, but the record in fact belongs to the class C_1, the record is classified as a false negative (fn). The confusion matrix then reports the number of records in the test set that belong to the different categories.

		Predicted Class	
		Class C_1	Class C_2
Actual Class	Class C_1	true positive (tp)	false negative (fn)
	Class C_2	false positive (fp)	true negative (tn)

Figure 3.4.: Confusion Matrix

Using this confusion matrix, multiple indicators have been proposed in order to assess the performance of a machine learning algorithm.

The following five indicators are considered as references and are used in the context of this thesis (Alpaydin, 2010; Witten *et al.*, 2011):

- **Success Rate**
  The success rate indicates the proportion of correctly classified records in the test set considering both classes C_1 and C_2.[14] The success rate is calculated as follows:

$$Success\ Rate = \frac{tp + tn}{tp + tn + fp + fn} \qquad (3.10)$$

- **Precision**
  The precision, also called accuracy, indicates to which extent the classification performed by the machine learning model corresponds to reality. It is calculated as the proportion of records of class C_1 among all records which have been classified as C_1 by the machine learning model:[15]

$$Precision = \frac{tp}{tp + fp} \qquad (3.11)$$

- **Recall**
  The recall measure indicates to which extent the machine learning model is capable of retrieving the items that actually belong to the class C_1. It is calculated as the proportion of items that have been classified as C_1 by the machine learning model among all items which actually are of class C_1:[16]

$$Recall = \frac{tp}{tp + fn} \qquad (3.12)$$

- **F-Measure**

---

[14]   Respectively, the error rate can be calculated as the proportion of incorrectly classified records

[15]   The precision can also be calculated for the class C_2. Its calculation is then: $\frac{tn}{tn + fn}$.

[16]   The recall can also be calculated for the class C_2. Its calculation is then: $\frac{tn}{tn + fp}$

The F-Measure is a combination of the precision and recall values and it is calculated as their harmonic mean (Witten *et al.*, 2011, p.175):

$$F = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{3.13}$$

- **Kappa Statistic**
  The Kappa Statistic compares the success rate obtained by a specific machine learning algorithm with the success rate achieved by a "random" algorithm that would randomly allocate the records in the test set to the class C_1 and C_2 with respect to the actual proportion of items of class C_1 and C_2 in the test set. The Kappa Statistic is then calculated as the performance increase between both success rates (Witten *et al.*, 2011, p.166).
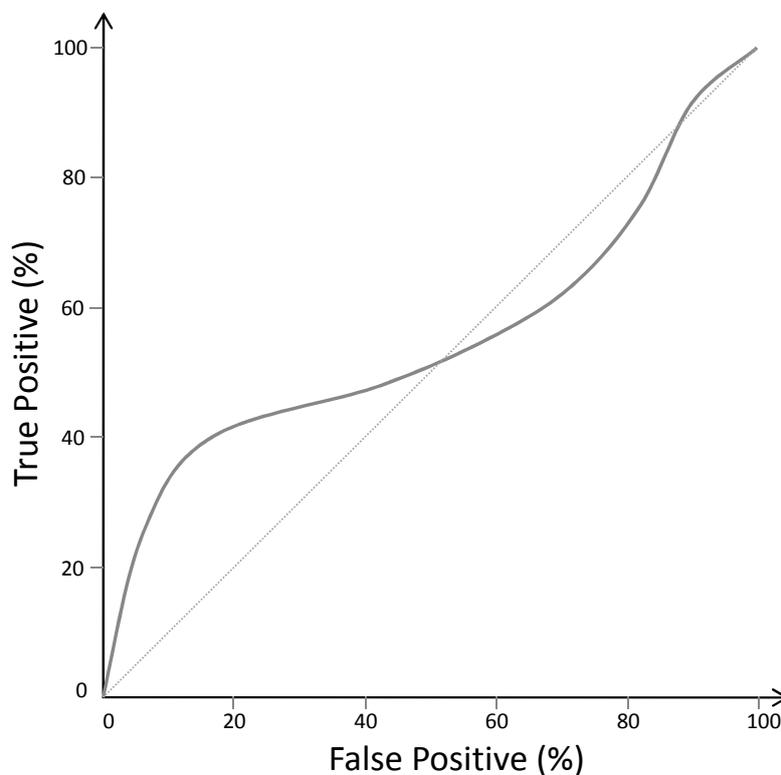
Figure 3.5.: ROC Curve

Finally, in order to graphically represent the performance of the different machine learning algorithms applied in chapter 7, the "ROC

curve" is used in this thesis. ROC means "Receiver Operating Characteristic." This graphical technique provides a representation of the true positive rate as a function of a the false positive rate, both presented as a percentage (Witten *et al.*, 2011, p.174). The ROC curve provides the ability to visualize the trade-off between these two parameters performed by different machine learning algorithms. For instance, the ROC curve illustrated in figure 3.5 shows that in order to achieve a true positive rate of 40%, the false positive rate will be equal to 10%. However, in order to achieve a true positive rate of 60%, the false positive rate will be much higher and equal to 65%. This means that this algorithm is efficient if the objective is to select samples with 40% of true positive records, but inefficient if the objective is to select samples with 60% of true positive records. The *x* and *y* axis on the ROC curve are calculated as follows:

$$x = \textit{False Positive Rate} = 100 \times \frac{fp}{fp + tn} \tag{3.14}$$

$$y = \textit{True Positive Rate} = 100 \times \frac{tp}{tp + fn} \tag{3.15}$$

Part II.
Conceptual Model

4. Customer Intimacy Breakdown Analysis

In chapter 2, the value discipline customer intimacy has been explained and put in relationship to other marketing concepts such as relationship marketing, key account management, and the service-dominant logic. The objective of this chapter is to establish how this concept can be broken down in multiple component parts, laying thereby the foundation of the overall model to assessing and monitoring customer intimacy. Based on an analysis of the definition of customer intimacy and of the constraints of the B2B context, various customer intimacy components have been determined at both organizational and individual levels. This chapter will develop this analysis, specify in detail each of these components, and motivate their relevance for the assessment of customer intimacy.

Section 4.1 will analyze existing approaches for assessing customer intimacy and will outline the distinctive characteristics of the proposed approach. Section 4.2 will subsequently elaborate on the performed customer intimacy breakdown analysis upon which the customer intimacy model proposed by this thesis is derived. This model consists of two parts, namely the *acquired* customer intimacy and the *leveraged* customer intimacy. Section 4.3 will detail the components

pertaining to the acquired customer intimacy and section 4.4 will develop the leveraged customer intimacy components.

4.1. Existing Approaches for Assessing Customer Intimacy

The measurement of customer intimacy has been a research topic addressed from multiple perspectives over the past years. In order to classify the different solutions proposed in existing literature, three criteria have been considered:

- **Analysis Level**
  Several degrees of analysis should be considered in order to thoroughly assess the degree of customer intimacy. While a general analysis of the activities involving both the customer and the provider at the organizational level is required, such as projects and sales contracts, a more detailed perspective focusing on the interactions occurring between the provider and customer employees is also needed to precisely estimate which employees and which teams in the provider organization have become "customer intimate". Thus, the customer intimacy assessment should be performed at the organizational level as well as at the individual level.

- **Assessment Focus**
  In this thesis, the objective is to assess the degree of customer intimacy established with different customers. The focus of the assessment is, therefore, on customers and more specifically on the interactions, activities, and projects involving the different customers. There are, however, other approaches to assess customer intimacy that take a different perspective and focus on the internal ability of a firm to implement a customer intimacy strategy. The assessment focus can therefore be on customers or on the provider organization itself.

- **Assessment Type**
  Two different approaches for measuring customer intimacy have

been investigated in past literature: the analytical approach which focuses on creating some key indicators out of existing data and the empirical approach which uses employees' feedbacks by means of questionnaires and interviews.

Different solutions to assess customer intimacy have been reviewed in the scope of this thesis. A selection of the most relevant ones as well as their categorization along the three criteria analysis level, assessment focus, and assessment type is provided in table 4.1. These solutions are detailed in the next paragraphs.

Cuganesan (2008) examines the use of financial data to calculate customer intimacy at the organizational level. Based on a case study with a wholesale financial service company, he suggests two modes of calculation which differs in the way customer intimacy is enacted: a "sales calculation network" approach and a "numeric calculation network" approach. The sales calculation network approach is driven by relationships, sales, and business units managers and focuses on the generation of knowledge about the interests of customers. The numeric calculation network approach is driven by the market intelligence department and focuses on the creation of performance measures based on market research. However, no details are provided on how these approaches are actually calculated.

In a balanced scorecard evaluation, Niven (2002) proposes five attributes which can be developed in order to measure customer intimacy. These are customer knowledge, offered solutions, penetration, culture of driving client success, and relationships in the long term. The operationalization and detailed implementation of these attributes, however, remain open.

Kaplan (2005, p.1) suggests that "for a differentiated customer intimacy strategy to succeed, the value created by the differentiation – measured by higher margins and higher sales volumes – has to exceed the cost of creating and delivering customized features and services." Later, he suggests to utilize the time driven activity based costing introduced in Kaplan & Anderson (2007) in order to assess these costs and evaluate customer profitability.

Table 4.1.: Overview of Existing Approaches Towards the Assessment of Customer Intimacy

	Analysis Level		Assessment Focus		Assessment Type	
Reference	Organization	Individual	Customer	Intern	Empirical	Analytical
Cuganesan (2008)	✓		✓			✓
Niven (2002)	✓		✓			✓
Kaplan (2005)	✓		✓			✓
Industry Directions & IBM (2006)	✓		✓			
Potgieter & Roodt (2004)	✓			✓	✓	
Tuominen *et al.* (2004)	✓		✓		✓	
Abraham (2006)	✓	✓	✓	✓		
Yim *et al.* (2008)	✓	✓	✓		✓	
This thesis	✓	✓	✓		✓	✓

An executive report suggests that services provide the opportunity for industrial companies to significantly deepen the level of customer intimacy and increase customer control, but it does not explain how to evaluate this level of customer intimacy and, thus, how to measure the improvement through the added services (Industry Directions & IBM, 2006).

Potgieter & Roodt (2004) provide a model in which they consider customer intimacy from the internal perspective and they conceive a questionnaire for the assessment of the customer intimacy culture of an organization. This questionnaire was validated by an empirical study in a company from the entertainment industry. Their approach does not consider the actual intimacy achieved with individual customers, but the ability of an organization, and more specifically its cultural aspects, to support a customer intimacy strategy.

Tuominen *et al.* (2004) provide a six-layer approach for evaluating customer intimacy: they differentiate whether the organization (1) was involved in the customer's planning process, (2) involved customers in their planning process, (3) partnered and jointly planned with customers, (4) aligned each other's operating processes, (5) designed operational interfaces, and (6) formalized the system of joint decision making. They use this scale to correlate the degree of customer intimacy with the degree of market orientation of the firm and its internal market intelligence capability, and recognize the importance of partnership and collaboration in the development of a customer intimacy strategy. However, only a few details are provided on actual implementation, and this solution solely focuses on the organizational level.

Abraham (2006, p.1) emphasizes the importance of the relationships between employees. He explains that customer intimacy represents "the formal or informal set of relationships established between supplier and customer, with a diverse array of partners, from corporate leadership to functional leadership (engineering, marketing, operations, maintenance, or service) and end-users of products or services." These dynamic relationships provide multiple points and frequency of contacts between the company and its customer, as well

as multiple points of view about the relationship and its benefits to both parties. According to his work, increasing customer intimacy can be achieved by improving the attitude of the employees dealing with the customer.

Yim *et al.* (2008) propose a model in which they consider both the customer-staff and customer-firm interactions in parallel. They define intimacy as the bondedness and connectedness of a relationship between two individuals and investigate how intimacy and passion can enrich customer service interactions and impact the customer-firm relationship. They validate this model by means of two empirical studies and conclude in particular that customer-staff affection influences customer-firm affection and customer-firm affection has a mediating role in strengthening customer loyalty.

This literature review outlines the distinctiveness of the approach proposed by this thesis. Indeed, as depicted in table 4.1, most of the existing solutions focus on the organizational level of analysis and do not consider the degree of customer intimacy established among employees. This thesis, on the contrary, considers both the organizational and the individual levels of analysis. Then, similarly to several other solutions, this thesis focuses on the actual assessment of the degree of customer intimacy established with different customers rather than on the inherent ability of an organization to pursue the customer intimacy value discipline. This thesis in addition combines an analytical customer intimacy measurement with an empirical assessment in order to validate the proposed solution.

4.2. Overview of the Customer Intimacy Breakdown Analysis

This section sets out the overall model to break down customer intimacy into multiple components. Many different aspects should be considered when developing a model to assess the degree of customer intimacy between a company and its customers. Liljander & Strandvik (1995) identified within their service relationship quality model that some of these aspects are at the organizational level,

while others are at the individual or employee level. Based on this premise, and in order to achieve the benefits outlined chapter 1, the model proposed by this thesis intends to include an assessment of the degree of the customer intimacy established with customers at both the organizational and individual levels. On the one hand, the individual level of analysis refers to an assessment focusing on customer and provider employees considered on an individual basis. On the other hand, the organizational level of analysis refers to an evaluation of the customer intimacy components considering the customer organization as a whole. The customer organization can be a team, a business unit, or the entire enterprise (see chapter 1, figure 1.1).

As developed in chapter 2, achieving customer intimacy does not solely consist of developing qualitative relationships with customers. Customer intimacy relates to the management of business relationships as well as to the management of customer related knowledge. More specifically, a successful customer intimacy strategy transforms these relationships and knowledge into competitive advantages. The decomposition of the concept of customer intimacy which is performed in this thesis in grounded on this analysis and roots in the original definition of customer intimacy presented by Treacy & Wiersema (1993, p.87): "to continually tailor and shape products and services to fit an increasingly fine definition of the customer." This definition can be split in two different parts: *acquired* customer intimacy and *leveraged* customer intimacy:

- The acquired customer intimacy refers to obtaining and understanding this "fine definition of the customer." It relates to establishing business relationships and obtaining customer related knowledge in order to determine means to adapt the value proposition to the specific needs of each customer.

- The leveraged customer intimacy concerns the actual competitive advantages achieved through business relationships and customer related knowledge. It represents the active part of the customer intimacy definition: "to tailor and shape products and services". These competitive advantages, such as customization and proactiveness are developed in section 4.4.

Leveraged Customer
Intimacy

Undirected Adaptability	Customer Intimacy
Standard Solution for Anonymous Markets	Inflexible Response to Customer Needs

Acquired Customer
Intimacy
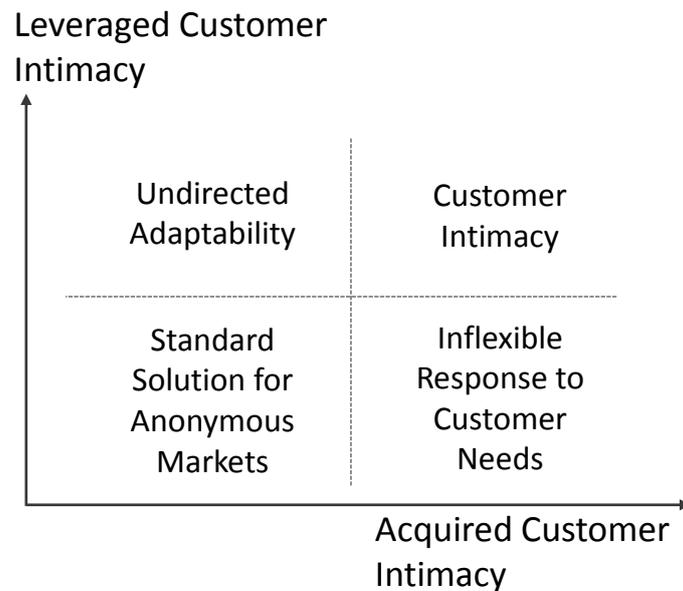
Figure 4.1.: The Two Dimensions of Customer Intimacy

As illustrated in figure 4.1, both the acquired and leveraged customer intimacy are required in order to effectively achieve a customer intimacy strategy:

- Considering the lower-left element of the quadrant which is defined as *standard solution for anonymous markets,* if the provider does not manage customer related knowledge and business relationships in order to obtain information on the specific customer requirements, nor try to individually adapt its solution to its customers, then this firm does not pursue customer intimacy by any means and should try to become a product leadership or operational excellence driven organization.

- The lower-right element – *inflexible response to customer needs* – describes companies that have established business relationships and effectively gathered customer related knowledge. These organizations, however, are unable to put these into actions in order to achieve a competitive advantage. For instance, if a relationship manager presents customer requirements to the provider organization, but the product development team rejects them and let the customer work with the standard offe-

ring, then the customer is left out with an inflexible response to its needs. This notion of "action on knowledge and relationships" reflects in part the definition of market orientation presented in section 2.3.2: "the organization-wide responsiveness to the generation and dissemination of market intelligence pertaining to current and future customer needs" (Jaworski & Kohli, 1993, p.54). Thus, in this configuration, the provider does not achieve a customer intimacy strategy with his customers.

- The upper-left element of this figure, called *undirected adaptability*, may be unrealistic. It refers to organizations which are not aware of this "fine definition of the customer": they do not have knowledge about the needs and expectations of their customers, nor business relationships to allow them to access this information. However, they build their value proposition around the creation of individualized solutions. Consequently, their offering can only by chance fit their customers' requirements and they also do not achieve customer intimacy.

- Finally, the upper-right element, which represents the actual customer intimacy strategy, refers to organizations which both obtain the fine definition of the customer and use it in order to generate a competitive advantage: they acquire a certain degree of customer intimacy with their customers and they are able to leverage it. Such organizations effectively manage both customer related knowledge and customer relationships. They also convert these two assets in a way that allows them to improve their value proposition and differentiate it from the standard ones proposed by their competitors.

In order to detail the requirements on organizations implementing a customer intimacy strategy, the sections 4.3 and 4.4 further break down the acquired and leveraged customer intimacy parts in multiple customer intimacy components. An overview of these components is proposed in figure 4.2.

```
┌─────────────────────────────┐   ┌─────────────────────────────┐
│   Acquired Customer Intimacy│   │  Leveraged Customer Intimacy│
│                             │   │                             │
│  Acquired    │  Acquired    │   │ Customization │ Customer     │
│  Knowledge   │  Knowledge   │   │               │ Loyalty      │
│  (Individuals)│ (Organization)│  │               │              │
│  ────────────┼────────────  │   │ Proactiveness │ Cross-selling│
│  Established │  Established │   │               │              │
│  Relationships│ Relationships│  │ Customer      │ Transaction  │
│  (Individuals)│ (Organization)│ │ Participation │ Costs        │
│                             │   │               │ Reduction    │
└─────────────────────────────┘   └─────────────────────────────┘
```
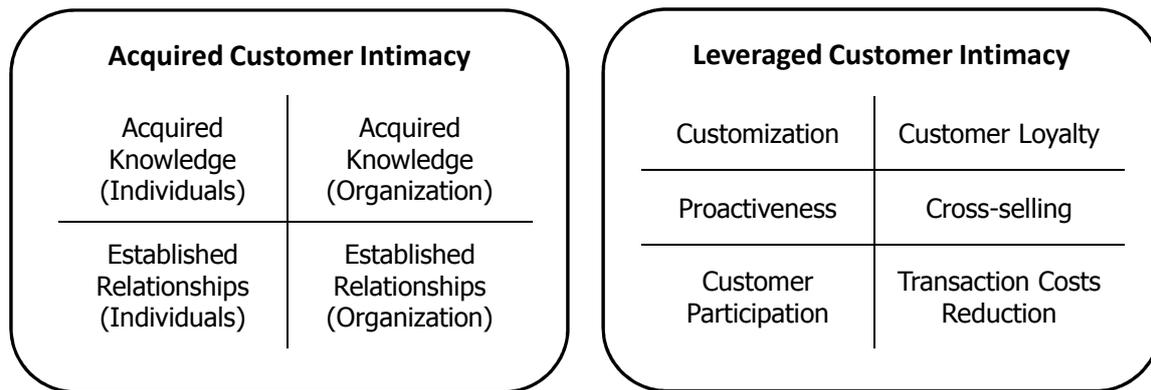
Figure 4.2.: Breakdown Analysis of the Acquired and Leveraged
            Customer Intimacy

4.3. Acquired Customer Intimacy Components

The concept of acquired customer intimacy has been created in this thesis in order to encompass the notion of "fine definition of the customer" presented in the previous section. As introduced in chapter 2, customer intimacy differentiates itself from other marketing strategies in the sense that it focuses on both customer related knowledge and customer relationships (see table 2.1). Customer related knowledge is required in order to understand the customer's current and future needs, as well as to determine which knowledge should be provided to the customer. Then, establishing qualitative customer relationships is necessary for a customer-intimacy driven organization as relationships are the means to become a reliable and trusted partner of the customer as well as to obtain further valuable knowledge and information from the customer that can be used to improve the value proposition. Manasco (2000, p.66) confirms that "relationships and knowledge are inseparable" in order to capitalize customer knowledge. Consequently, the two components pertaining to the acquired customer intimacy are:

- Acquired customer knowledge
- Established customer relationships

As previously explained, a requirement in the approach followed by this thesis is to perform the customer intimacy assessment at two

different levels of analysis: the individual and the organizational levels. Indeed, in order to accurately understand the customer, it is necessary to identify the provider employees who have knowledge of, and relationships with, the overall organization, as well as those who have knowledge of, and relationships with, specific employees inside the customer organization. Therefore, in this thesis, the two customer intimacy components acquired customer knowledge and established customer relationships are assessed at both the individual and organizational levels, as depicted in figure 4.2. These components are further detailed in the next two parts of this section.

4.3.1. Acquired Customer Knowledge

The first component of the acquired customer intimacy refers to the acquisition and development of customer knowledge. Batt (2004, p.172) explains that in order to achieve customer intimacy, "the firm must keep deepening its knowledge of the customer and put this knowledge to work through the organization." As a matter of fact, customer intimacy requires advanced knowledge management capabilities. Zack *et al.* (2009) confirms that the organizations that pursue the value discipline customer intimacy have implemented the widest range of knowledge management practices. Moreover, a positive correlation has been established between customer knowledge development and service activities as "service relationships offer an opportunity for greater customer knowledge to be developed by the employees because of their repeated interactions with the same customer" (Gwinner *et al.*, 2005, p.136). In a product development context, customer knowledge development has been defined as "a process of developing an understanding of customer new product preferences that unfolds through the iteration of probing and learning activities" (Joshi & Sharma, 2004, p.48). Taking a broader perspective, Bueren *et al.* (2004) distinguish three categories of customer knowledge: *about*, *for* and *from* the customer.

Knowledge *about* the customer is certainly the most important one to develop a customer intimacy strategy.

- At the organizational level, knowledge *about* the customer refers to gaining an understanding of the current and future needs of the customer, to obtaining information about the customer strategy and about its mid- and long-term development. It also includes knowledge about the interaction history with the customers such as the projects performed with the customer and the products and services purchased by the customer. Knowledge *about* the customer also consists of the inherent description of the customer organization which provides valuable information to optimize the interaction with the customer, such as the organizational structure, the customer's behavior and its purchasing process. While some of this knowledge is certainly explicit, such as the description of prior projects, opportunities and contracts, a part of this knowledge is also implicit. For instance, a project manager who completed successful projects with the customer most likely gathered information about its future needs and planned developments while a key account manager is aware of its customer's purchasing processes.

- At the individual level, knowledge *about* the customer refers to knowledge about customer employees, such as specific needs, preferences, and behavior. This aspect is particularly important in a B2B context as the customer consists of multiple stakeholders, such as the users and buyers, which all have different requirements. In order to be successful and to optimize its value proposition, the provider must be able to manage all these different expectations (Homburg & Jensen, 2004). Gibbert *et al.* (2002, p.3) argue that "smart companies [...] seek knowledge through direct interaction with customers, in addition to seeking knowledge about customers from their sales representatives". For instance, some customer employees may need a specific service level agreement because they use a service differently from the rest of the organization.

Knowledge *for* the customer aims at fulfilling the customer's needs with regard to his knowledge requirements. It refers essentially to information about the value proposition such as technical details on the purchased products and services. This category of knowledge

also includes insight into the customer's industry which might be relevant for the customer in order to generate future needs in the customer organization such as new regulations, or new market opportunities. The consideration of multiple level of granularity, from the entire organization perspective, down to the teams and the individuals perspective is also required as different customer teams and customer employees will have different requirements in terms of knowledge: depending on their role, they will expect business, technical, or financial information.

Finally, knowledge *from* the customer consists of the information related to the products and services of the provider that the customer employees acquire by using them. This includes information such as the quality, reliability, or usability of the products and services. This knowledge also includes information on the satisfaction of the customer as well as suggestions from the customer for new products or service developments. If provider employees are able to access this knowledge and to convey it back in their organization, this knowledge from the customer becomes a highly relevant asset for adapting and improving the value proposition.

4.3.2. Established Customer Relationships

The second component of the acquired customer intimacy consists of the relationships established between the provider and the customer, at both individual and organizational levels. Relationships are an inherent part of any business ecosystem and become steadily more intensive in the current globalized economy (Donaldson & O'Toole, 2007). They have become an increasingly important matter of study in marketing literature, as several analyses demonstrate their positive influence on business performance (Narver & Slater, 1990; Varadarajan & Rajaratnam, 1986; Reichheld & Sasser, 1990).[1] It has been explained in chapter 2 that the value discipline customer intimacy is grounded in the concept of relationship marketing and relies on the establishment of business relationships. In short, Donaldson

---

[1]   Further details are provided in section 2.2.1.2.

& O'Toole (2007, p.13) summarize the benefits derived from established business relationships in order to support a customer intimacy driven strategy: business relationships help "identifying customer needs and requirements, anticipating future trends and monitoring environmental forces, and satisfying customers' existing and future requirements."

Business relationships have been assessed in multiple ways over the past decades, and several studies which evaluate the constituents of a relationship in a commercial setting are already available (Morgan & Hunt, 1994; Odekerken-Schröder *et al.*, 2003; Bove & Johnson, 2001; Barnes, 1997). In previous literature, the assessment of customer relationship is referred to as relationship quality or relationship strength. Even though Richard (2008) argues that much literature uses these two terms equally, Bove & Johnson (2001, p.190, p.193) propose to distinguish the two concepts. They define relationship quality as "an overall construct which is based on all previous experiences and impressions the customer has had with the service provider", and relationship strength "as the magnitude of a relationship between two individuals in a commercial setting." In this perspective, relationship quality is more focused on the organizational level while relationship strength concerns predominantly the individual level. In past literature, the most often cited characteristics of relationship quality and relationship strength are trust and commitment[2] (Richards & Jones, 2008; Roberts *et al.*, 2003; Lages *et al.*, 2005). Therefore, assessing established relationships refers to understanding the degrees of trust and commitment established between the provider, the customer, and their respective employees:

- Trust has been conceptualized as having "confidence in an exchange partner's reliability and integrity" (Morgan & Hunt, 1994, p.23). It was further refined along the following three dimensions: contractual trust, goodwill trust, and competence trust (Sako, 1992). Contractual trust is determined by the re-

---

[2]   Communication quality, customer satisfaction, social bonds, and information flows are further aspects that have been identified as characteristics of relationship quality and strength.

spective legal obligations of both partners. Goodwill trust refers to a mutual commitment and support to each other, including confidence that the partners will not try to take an unfair advantage of each other. Finally, competence trust has been defined as the belief that the partner has the ability, technical knowledge, expertise, and capability to perform his role (Sako, 1992).

- Commitment was defined by Anderson & Weitz (1992, p.19) as "a desire to develop a stable relationship, a willingness to make short-term sacrifices to maintain the relationship, and a confidence in the stability of the relationship." This translates in the provider organization and its employees into a readiness to help the customer solving his problems, into demonstrating an adequate flexibility when needed by the customer, and into seeking the best solution from the customer's perspective on the long-term rather than from the provider's perspective on the short term.

Considering the individual level of analysis, acquired customer knowledge and established customer relationships are intricately connected. Ballantyne (2004, p.119) introduces the concept of relationship specific knowledge which he considers as a mediator for the development of trust and for the generation of business knowledge. He argues that this is a "kind of tacit knowledge that might have positive use in dealing with current dilemmas and determining future expectations." Reciprocally, Gummesson (2008, p.190) establishes the knowledge relationship as the 21st of his 30 "R" of relationship marketing. He argues that knowledge is "not only embedded in an individual, group, or corporation, but also in the relationships between companies." This knowledge relationship builds upon a complex network of social ties established between provider and customer employees and it is referred to as a social structure (Donaldson & O'Toole, 2007, p.116). This social structure, when used appropriately, is highly valuable for the provider as it can become a strategic lever in order to improve the value proposition and to develop the customer intimacy strategy (Dalkir, 2011, p.170). This potential value of the social structure is called social capital. Dalkir (2011, p.474)

defines social capital has "the value created when a community or society collaborates and cooperates (through such mechanisms as networks) to achieve mutual values." In this thesis, the assessment of the established relationships at the individual level corresponds to the assessment by means of social network analysis of the social structure established between provider and customer employees.

In the next section of this chapter, the components pertaining to the leveraged customer intimacy will be introduced.

4.4. Leveraged Customer Intimacy Components

The second part of the customer intimacy breakdown analysis is called *leveraged* customer intimacy. While the acquired customer intimacy concerns the investments made by the provider in order to obtain some knowledge of, and to establish some relationships with, the customer, the leveraged customer intimacy refers to the actual competitive advantages, benefits, and value proposition improvements achieved by the provider by "leveraging" these knowledge and relationships. When the provider uses his knowledge of, and relationships with, the customer, he adapts, transforms, or enriches his offering to the customer, thereby improving his value proposition, and convincing the customer to choose him as a provider rather than other competitors.

In order to fully understand the leveraged customer intimacy, a thorough review and analysis of literature has been performed in this thesis. This analysis has led to decompose the leveraged customer intimacy into the following six components, as depicted in figure 4.2: customization, customer loyalty, proactiveness, cross-selling, customer participation, and transaction cost reduction. The next parts of this section elaborate on each of these components, outline their association with the value discipline customer intimacy, and demonstrate why they lead to the generation of competitive advantages and benefits for the provider. The actual metrics created in this thesis for assessing these six components upon existing customer data will be introduced in chapter 5.

4.4.1. Customization

Customization is the first component of the leveraged customer intimacy part of the model proposed by this thesis. Customer intimacy driven organizations, with their objective to "tailor and shape products and services to fit an increasingly fine definition of the customer" (Treacy & Wiersema, 1993, p.87) inherently rely on customization strategies which "aim at providing customers with individually tailored products and services" (Gwinner *et al.*, 2005, p.131). Customization is particularly important in the B2B context because it is closely related to the servitization process that has occurred over the past decades. Servitization, which refers to a business model shift from selling products to selling "customer-focused combinations of goods, services, support, self-service and knowledge" is, as a matter of fact, a form of customization (Vandermerwe, 1988, p.314).

Several analyses in past literature have confirmed the importance of customization in order to create a competitive advantage and to improve the value proposition. Fornell *et al.* (1996, p.8) demonstrated with their American customer satisfaction index that customization, which they defined as "the degree to which the firm's offering is customized to fit heterogeneous customer needs" has a more significant impact on customer satisfaction than reliability. Richards & Jones (2008, p.126), in an analysis aiming at finding the value drivers of customer relationship management, observed that "increased customization of products and services is positively related to brand equity and relationship equity in the maintenance stage." Thus, customization increases the provider's value from the customer's perspective. Finally, Vargo & Lusch (2004b, p.326) confirmed the importance of customization in contrast to standardization as they state that "the normative marketing goal should be customization, rather than standardization." They thereby indicate that if standardization increases production efficiency, it also decreases marketing effectiveness: the heterogeneity of the customer demand requires individually tailored response that standard offerings are unable to provide. Therefore, organizations should consider customization rather than standardization as their primary marketing focus.

In order to develop a customization strategy, different approaches have been proposed, in particular, mass customization, customerization, and service customization through employee adaptiveness. An analysis of these different concepts leads to the conclusion that customization, within the scope of this thesis, is aligned with the service customization through employee adaptiveness approach:

- **Mass Customization**
  Mass customization can be perceived as a means to combine standardization with customization, thereby achieving both cost efficiency and marketing effectiveness. It is defined as: "a system that uses information technology, flexible processes, and organizational structures to deliver a wide range of products and services that meet specific needs of individual customers (often defined by a series of options), at a cost near that of mass-produced items (Silveira *et al.*, 2001, p.2). Mass customization does not fit into the model proposed by this thesis as it leverages information technology rather than acquired knowledge of, and established relationships with, customers in order to achieve customization.

- **Customerization**
  Customerization has been proposed as an evolution of mass customization which gives more controls to customers in the design of products and services, and relies on interactions with customers to achieve customization (Wind & Rangaswamy, 2001). Through an emphasis on knowledge *from* customers as well as a redefinition of the role of the customer as an active co-creator, customerization is, to some extent, close to customization as proposed by this thesis. Customerization, however, does not consider the development of interpersonal relationships with the customer, but, similarly to mass customization, leverages IT systems in which customers directly input their requests and preferences in order to provide customers with customized solutions. Wind & Rangaswamy (2001, p.15) confirm that "mass customization is IT-intensive on the production side, whereas customerization is IT-intensive on the marketing side." Thus, customerization and customization in the context of this the-

sis are different concepts, even though some similarities can be observed.

- **Service Customization through Employee Adaptiveness**
  While mass customization and customerization intend to achieve customization mainly though the use of information technology, Gwinner *et al.* (2005) emphasize the importance of the provider employees in order to achieve customization in their *service customization through employee adaptiveness* model. They argue that customer knowledge is antecedent to effective customized service behaviors. However, in contrast to mass customization and customerization, customer knowledge is not generated by information systems but resides in the front-line employees who have regular interactions with the customer. Thus, this model corresponds to the approach proposed by thesis: the objective is to leverage the acquired knowledge of, and the established relationships with, customers to thoroughly understand the customers' explicit and tacit needs. These assets are used to customize the offering and, thus, to achieve a competitive advantage. Gwinner *et al.* (2005, p.136) confirm as a matter of fact the importance of interpersonal relationships for service customization as they argue that "service relationships offer an opportunity for greater customer knowledge to be developed by the employees because of their repeated interactions with the same customer."

4.4.2. Loyalty

Customer loyalty is the second leveraged customer intimacy component. In its definition of customer loyalty, Oliver (1999, p.34) insists on the establishment of a stable and long-term relationship between the provider and the customer: customer loyalty is "a deeply held commitment to rebuy or repatronize a preferred product/service consistently in the future, thereby causing repetitive same-brand or same brand-set purchasing, despite situational influences and marketing efforts having the potential to cause switching behavior." According to Treacy & Wiersema (1997, p.40), customer loyalty is the

most important benefit derived from a customer intimacy driven strategy: "the customer-intimate company's greatest asset is, not surprisingly, its customers' loyalty."

Multiple analyzes have corroborated the importance of customer loyalty over the past decades, since Reichheld & Sasser (1990) established that a 5% improvement in customer retention can lead to a 25% to 85% profitability improvement. This finding created a strong impulse for research that analyzes the relationship between customer loyalty, customer retention, and customer satisfaction (Dick & Basu, 1994). Reichheld & Teal (2001, p.39) elaborated five benefits and competitive advantages which are derived from customer loyalty. These are the reduction of the customer acquisition costs, the per-customer revenue growth, the operating costs reduction, the generation of referrals and recommendations, and the payment of price premiums by loyal customers. Even though an empirical analysis in the context of B2C financial services found that loyalty is not positively associated with profitability (Storbacka et al., 1994), Grönroos (2007, p.8) and Heskett et al. (1994) confirmed that loyal customers are in most cases profitable.

It has been widely recognized in past literature that well established relationships are an antecedent to customer loyalty, thereby linking customer loyalty to the acquired customer intimacy part of the model proposed by this thesis. For instance, Hennig-Thurau et al. (2002) established that the key aspects trust and commitments of relationship quality directly or indirectly impact the customer's loyalty. Palmatier et al. (2007), in an analysis at both the organizational level and at the employee level argued that relationship-enhancing activities, such as actions and efforts that strengthen relationship quality positively influence both the loyalty to the sales persons and to the provider organization. Finally, Ndubisi (2007) empirically proved that relationship marketing endeavors are positively correlated with an augmentation of the degree of customer loyalty.

4.4.3. Proactiveness

The breakdown analysis of the leveraged customer intimacy part of this model has led to define proactiveness as the third leveraged customer intimacy component. With an emphasis on customers and on customer needs, Sandberg (2007, p.253) defines customer-related proactiveness as "acting based on the information gathered about the customers before their behavior has had a direct impact on the firm, or deliberately influencing and creating changes in customer behavior." This definition outlines the importance of acquiring customer-related knowledge and insight in the customer's industry, and of using this knowledge as a trigger of the customer related activities. Thus, in a customer related proactiveness configuration, the provider initiates the interaction process with the customer instead of awaiting its explicitly articulated demands. In a similar way, Treacy & Wiersema (1997, p.127) confirm the importance of customer-related proactiveness for successful customer-intimate organizations when they argue that "a customer intimate firm uses its superior expertise in the client underlying problem to change the way the customer does business."

Proactiveness is often contrasted with reactiveness which indicates a focus on understanding and fulfilling customer requirements, thereby reacting to customer behavior (Sandberg, 2007). Thus, customer-intimate organizations combine both reactiveness and proactiveness. They are reactive as they work towards fulfilling to the highest extent the customer needs. They are also proactive as they try to transform and shape the customers operations, structures, and behavior in order to solve his problems, even before the customer is able to request for it.

Different types of proactiveness have been investigated in previous literature, in particular the proactive service improvement and the proactive service recovery:

- Wallenburg (2009, p.78) focuses on proactive service improvement in a B2B context and brings proactiveness in the context of innovation. Considering an innovation which is potentially

beneficial to the customer, either by leading to a cost reduction or a performance improvement, a proactive improvements occurs if the provider "proactively enhances the service provided to that specific customer" with this innovation without the customer asking for it. Wallenburg (2009) establishes that both types of performance improvements are strong drivers of customer loyalty, thereby supporting the customer intimacy strategy.

- De Jong & De Ruyter (2004, p.458) elaborate on the importance of adaptive and proactive behavior in service recovery. Adaptive behavior refers to the actions undertaken by employees in response to specific customer problems whereas proactive behavior concerns problem-independent customer-related activities such as "soliciting suggestions from customers, detecting and correcting causes of service problems and challenging existing routines." De Jong & De Ruyter (2004) argue by means of an empirical analysis that while adaptive behaviors positively influence the customer's degree of loyalty, proactive behaviors lead to additional service revenues.

4.4.4. Cross-selling

The fourth leveraged customer intimacy component refers to the cross-selling achievements of the provider. Kamakura *et al.* (1991) explain that cross-selling aims at increasing the number of different products and services sold to the customer and propose a predictive model to assess the likelihood of the customer to accept cross-selling driven offerings. Malms & Schmitz (2011, p.255) suggest a customer intimacy aligned definition of cross-selling: "an offer of customized solutions or the provision of a full assortment of products and services." Reciprocally, taking the customer's perspective, Venkatesan & Kumar (2004, p.111) define cross-buying as "the number of different product categories a customer has purchased." They establish this factor as a key element of their customer lifetime value assessment model and prove that it increase the customer's purchase frequency, thereby generating additional revenues. In order to achieve cross-selling, the provider should try to complement the original product

or service sold to the customer with other components that improve the overall solution delivered to the customer.

Looking at the relationship between customer intimacy and cross-selling, Akura & Srinivasan (2005, p.1008) demonstrate that customer intimacy and cross-selling are intricately connected and argue that firms "achieve customer intimacy when committing against a certain level of cross-selling." Treacy & Wiersema (1997) confirm that customer-intimate organizations inherently provide their customers with cross-selling offering as they do not only sell products but solutions combining multiple products and services that fulfill the exact customer's needs. In the B2B context, Harding (2004) recognizes the importance of cross-selling, but also argues that cross-selling can damage the relationship if performed with the objective to increase the provider's revenues rather than to provide the customer with the solution that fits its requirements and solves its problems. He thereby links cross-selling with the component "acquired customer knowledge" of the acquired customer intimacy part of this model and confirms that deep customer knowledge is a prerequisite to effective cross-selling. This relationship between customer knowledge and cross-selling has also been confirmed by Akura & Srinivasan (2005, p.1007) who argue that "successful cross-selling requires customer intimacy and detailed information on customer preferences."

Achieving cross-selling leads to multiple benefits for the provider. In addition to the positive impact on revenues established by Venkatesan & Kumar (2004), cross-selling also improves customer's profitability as the costs to acquire the customer can be distributed on products and services of different categories. These costs are also reduced for any subsequent component added to the solution provided to the customer. Cross-selling also has an indirect impact on the customer loyalty as it increases the customer's switching costs and the customer retention rates (Kamakura *et al.*, 2003). If the customer purchased different products and services from the same provider, the costs of replacing all these components by other alternatives is higher than if he only bought one single product or service. Thus, an heterogeneous solution composed of multiple products and services is a motivating factor for the customer to remain loyal to its provider.

Finally, it has also been established that cross-selling increases the customer-related knowledge acquired by the provider (Kamakura *et al.*, 2003), thereby having a positive influence on acquired customer intimacy. Indeed, the variety of products and services purchased by the customer allows the provider to obtain a broader understanding of the customer needs and preferences.

4.4.5. Customer Participation

Customer participation is the fifth component of the leveraged customer intimacy. It has been defined as "the customer behaviors related to specification and delivery of a service" (Cermak & File, 1994, p.2). This aspect is fundamental in the previously described service-dominant logic which outlines that both the provider and the customer are co-creators: the customer is not a sole receiver of the value distributed by the provider, but actively participates in its creation by making his knowledge available to the provider (Vargo & Lusch, 2004a). Treacy & Wiersema (1997, p.136) confirm the importance of customer participation for the success of a customer intimacy driven strategy as they argue that "customer-intimate firms use their client to stay at the edge of new thinking". They quote an executive officer in a customer-intimate organization arguing that "the product is conceived at the customer's office" (Treacy & Wiersema, 1993, p.41).

Bettencourt (1997) identifies three different types of customer participation in his customer voluntary performance model:

- First, the customer can promote the provider organization and its offering into its network. This kind of participation indicates, as previously explained, the degree of loyalty of the customer. However, it does not lead to a co-creation of the value between the provider and the customer: the provider creates the offering without the customer.

- Secondly, cooperation can be another form of customer participation: the customer supports the service employees to achieve the expected service level agreements during the delivery phases, but the knowledge of the customer is not used in order to support the design of the provided solution.

- The third type of customer participation is in line with the approach of this thesis and refers to customers who act as "organizational consultants". Such customers actively participate in the design and implementation of the solution by making available their understanding and knowledge of the problem to be solved as well as by making recommendations for improving the provided solution. In that regard, Bettencourt (1997) argue that customers are a unique source of advice with an outstanding experience of the provider's products and services.

Satzger & Neus (2010, p.230) emphasize the importance of customer participation to support the provider's innovations in their C^4 framework. They suggest that customers are the most important source of service innovation and argue that "the most efficient place for service innovation may today lie outside of service provider organizations, i.e. within peer-networks of users who are intrinsically motivated to support innovation." Similarly, Magnusson *et al.* (2003) empirically compared innovations achieved by users and professional designers. Their finding was that users provided more original and user-focused innovations while professional provided innovation that are easier to implement. Consequently, organizations having their customers participating to the value creation process hold an important means to improve their value propositions and to achieve a competitive advantage.

Increasing customer participation in solution development also provides more structural benefits to the organization. Chesbrough (2007) argues that open business models involving customers lead to a reduction of the research and development costs and, thus, to an increase in profitability. Customer participation also allows organizations to obtain qualitative market intelligence data and to better target the marketing strategy to customers and prospective customers (Ndubisi, 2007). Finally, Cermak & File (1994) established that customer participation strengthens the relationship between the customer and the provider as well as increases the customer satisfaction.

4.4.6. Transaction Costs Reduction

The sixth component of the leveraged customer intimacy refers to a reduction of the transaction costs for organizations achieving customer intimacy. This notion of transaction costs has first been introduced by Coase (1937) as the costs of using the price mechanism in the market. He established that the actual costs of the customer for acquiring a product or a service in the market not only include the price paid by the customer to the provider, but also information costs, negotiation costs, and policy and enforcement costs. Similar costs are borne by the provider for selling his product and services since he has to inform the customer about its offering, negotiate the offer, and ensure that the offering fulfills the customer's expectations. According to Dyer & Chu (2003), these additional costs are strongly influenced by the existence of relationships and the establishment of trust among the provider and the customer. For instance, the information costs are reduced if the customer chooses not to invest time and resources to find a provider but simply select its preferred partner. The negotiation will run more smoothly between partners who already know each other. The customer's enforcement costs will be reduced if the customer trusts the provider in his ability to deliver the expected product or service, as fewer safeguards have to be setup. Dyer & Chu (2003, p.57) empirically proved that "trustworthiness lowers transaction costs and may be an important source of competitive advantage."

Williamson (1979) outlined the importance of transaction costs in the study of economics and defined three aspects impacting transaction costs: the transaction frequency, its uncertainty, and its idiosyncrasy, which reflects the uniqueness and individualization of the investments performed by the provider and the customer, such as the purchasing of special equipment by the provider in order to fulfill the contract. Therefore, customer intimate organizations which make some special efforts in term of time and cost investment in order to fulfill the customer requirements increase the idiosyncratic degree of the relationship and, thus, lower the transaction costs. Dyer (1997) analyzed the influence of relationship-specific investments on trans-

action costs and confirmed that such investments do not lead to an increase of the transaction costs and in some cases even lower them.

Focusing on the impact of customer loyalty, which is, as previously described, customer-intimate organizations' main asset, Reichheld & Teal (2001, p.39) established that qualitative relationships with loyal customers lead to a reduction of the acquisition and operating costs:[3]

- Several studies acknowledge that the acquisition of a new customer is significantly more expensive than the investments required to keep an existing customer (Grönroos, 2007, p.145). Thus, by focusing on their most important and most loyal customers, customer intimate organizations have the means to lower the acquisition costs. Treacy & Wiersema (1997, p.139) confirm that customer intimate companies should avoid "business that might generate only short-term revenues," and whose acquisition costs cannot be balanced with regular and long-term revenues.

- Operating costs are reduced in long-lasting relationships with loyal customers because the frequent and regular interactions between the provider and the customer lead to the creation of a common knowledge base between both organizations (Ballantyne, 2004). Thus, projects run in a smoother way as the provider better understands the customer's expectations and the customer can better articulate his requirements. In addition, in the context of repeatedly delivered services, fewer mistakes occur as service are performed more often, which in turn leads to an additional reduction of the operating costs (Grönroos, 2007, p.146).

Zajac & Olsen (1993) and Den Butter (2010) consider the transaction costs in the broader concept of value creation, and emphasize the notions of transaction value. Zajac & Olsen (1993) argue that the focus on single party transaction cost optimization should be replaced by a focus on transaction value and an emphasis on "joint value maxi-

---

[3] Other economic effects established by Reichheld & Teal (2001) include revenue growths, payment of price premiums, and referrals.

mization". This transaction value takes into account the interdependencies between the provider and the customer. Den Butter (2010, p.2) defines transaction management as "the ability to keep the costs of trade transactions as low as possible so that the value creation from these transactions is optimized."

In summary, this chapter demonstrated by means of a thorough literature review that the concept of customer intimacy can be broken down into two parts, namely the acquired customer intimacy and the leveraged customer intimacy. The acquired customer intimacy consists of the acquired customer knowledge and the established customer relationships. The leveraged customer intimacy consists of six components which are customization, loyalty, proactiveness, cross-selling, customer participation, and transaction costs reduction. These customer intimacy components have been thoroughly described and their association to the concept of customer intimacy has been motivated upon past literature. This analysis is foundational for the remaining of this thesis. In chapter 5, it will be explained how this thesis proposes to evaluate these components in an analytical manner, thereby achieving the overall objective to assess the degree of customer intimacy established with different customers as well as its impact on business.

5. *CI Analytics* Model and Methodology

In chapter 4, the concept of customer intimacy has been broken down in multiple components which pertain either to the acquired or to the leveraged customer intimacy. The objective of this chapter is to introduce the *CI Analytics* model to assess and monitor these components as well as to detail the *CI Analytics* methodology to calibrate and utilize the model. As it will be explained in the next sections, these model and methodology use social network analysis and datamining techniques, and leverage customer related data available in the provider's information system. An essential benefit of this approach is that the assessment of the customer intimacy components is performed automatically, once the calibration has been performed. Thus, in line with business intelligence and analytics systems, this thesis provides the ability to monitor the evolution of the customer intimacy components values over time in a continuous manner.

Section 5.1 presents an overview of the *CI Analytics* model and methodology. Sections 5.2 and 5.3 subsequently detail the relevant sources of customer intimacy data and the conceived metrics for assessing the acquired and leveraged customer intimacy components.

5.1. *CI Analytics* Overview

While part 5.1.1 elaborates on the *CI Analytics* methodology, part 5.1.2 details the key aspects of the *CI Analytics* model.

5.1.1. *CI Analytics* Methodology

As described in chapters 2 and 4, the choice to follow the value discipline customer intimacy impacts and even determines the provider's strategy and operational model. A well-driven customer intimacy strategy can be recognized along certain characteristics such as the evolution of customer relationships into longer term partnerships, the access to customers' information systems, some regularity in the interactions, the successful completion of joint activities with customers and the mutual involvement of top level management in these activities. The provider's information system contains elements of evidence for most of these characteristics. For instance, successful joint activities can be tracked in the project database. The interaction regularity as well as the involvement of top-level management can be assessed with an analysis of the different communication channels. The development of a partnership can be derived from the information contained in the customer relationship management system.

The *CI Analytics* approach aims to identify the relevant elements of evidence of the customer intimacy components inside the provider's information system as well as to define a means to aggregate them into understandable customer intimacy metrics in order to assess the degree of customer intimacy with each customer and at multiple levels of details. This approach poses three significant and interrelated challenges which are solved by the *CI Analytics* methodology. These three challenges are the following:

- **Inference Challenge**
  The inference challenge concerns the fact that the customer intimacy components are not directly observable and measurable inside the provider organization or at the interface between the provider and the customer (De Choudhury *et al.*, 2010). On the contrary to physical characteristics such as size or volume, or

even to explicit performance metrics such as revenue or profitability, concepts such as established relationships or acquired knowledge cannot be directly measured and, thus, must be inferred out of observable and available data, such as interactions, projects, and revenue records.

- **Relevance Challenge**
  The relevance problem relates to the fact that there is no exact specification of the data which is necessary and how it should be transformed in order to precisely infer each of the customer intimacy components (De Choudhury *et al.*, 2010). Indeed, the available data inside the provider's information systems can be combined in an infinite number of customer intimacy metrics, by simply varying the ways the different data items are aggregated. Thus, a key challenge is to identify the most relevant sources of customer intimacy evidence as well as the best metrics which reflect the actual values of the customer intimacy components. Moreover, a means to identify the actual values of the different customer intimacy components must be determined in order to validate the proposed approach.

- **Calibration Challenge**
  The third challenge makes the first two problems, the inference and the relevance issues, even more complex. This issue roots in the fact that each provider organization has its own way of interacting with its customers and manages customer related data in a specific manner. For instance, some providers prefer email communication while others prefer phone calls or face to face meetings. A three-months project may be considered as long in some organizations and as short in others. An organization may save all details about all interactions and activities with a customer, while another one keeps only the most relevant data. Thus, some metrics which are relevant for a specific organization might become less significant for another provider. Consequently, the generic customer intimacy metrics which have been conceived must be adapted and weighted by means of a calibration to the individual characteristics of each

provider organization, such as the interaction, activity, and data storage patterns.

	Step	Output
1	Break down the concept of customer intimacy into mulitple components	Customer Intimacy Breakdown Analysis
2	Identify the sources of evidence to assess the customer intimacy components	Customer Intimacy Evidence Sources
3	Define the customer intimacy metrics	Unweighted Set of Metrics
4	Calculate the customer intimacy metrics	Database Results (Analytical)
5	Empirically estimate the customer intimacy components	Empirical Results (Questionnaire)
6	Calibrate the model by applying data-mining techniques	Machine-Learning Models
7	Validate and interpret the model	Weighted Set of Metrics

Individually performed for each provider (steps 4–7)

Figure 5.1.: *CI Analytics* Methodology

The *CI Analytics* methodology intends to solve the inference, relevance, and calibration challenges, thereby adapting the model to the specific data and interaction patterns of each provider. Since this methodology relies on the analysis of customer related data in the provider's information system, its design is aligned to the knowledge discovery in databases process proposed by Fayyad *et al.* (1996b) and presented in section 3.2.1. The seven steps of the *CI Analytics* methodology are depicted in figure 5.1. They are detailed and put in relationship with the steps of the knowledge discovery in databases process in the next paragraphs. While the first three steps are generic and performed once, steps 4 to 7 aim at solving the calibration challenge and, thus, are individually performed by each provider.

1. **Break down the concept of customer intimacy into customer intimacy components**

   The first step of the *CI Analytics* methodology refers to a thorough analysis of the concept of customer intimacy and its breakdown analysis into multiple assessable components. This analysis represents an important contribution of this thesis and is elaborated in chapter 4. It establishes that customer intimacy can be broken down into two parts, the acquired and the leveraged customer intimacy. The acquired customer intimacy consists of two components: acquired customer knowledge and established customer relationships. The leveraged customer intimacy consists of six components which are customization, customer loyalty, proactiveness, cross-selling, customer participation and transaction cost reduction.

2. **Identify the sources of evidence to assess the customer intimacy components**

   The next step of the *CI Analytics* methodology is concerned with the identification of the relevant sources of evidence which can be analyzed to infer the customer intimacy components. A fundamental idea of the approach followed by this thesis is that the degree of customer intimacy established between a provider and customer is reflected to some extent in the provider's information systems. In order to determine these relevant sources of customer intimacy evidence, this thesis relies on previous research and past literature in the field of relationship marketing, customer relationship management, and social network analysis. The layer Customer Intimacy Data of the *CI Analytics* model presented in figure 5.2 outlines the multiple sources of customer intimacy evidence considered in the scope of this thesis.

3. **Define the customer intimacy metrics to calculate customer intimacy components out of the customer intimacy data**

   Closely related to the second step, the third step of this methodology consists of the actual design of the metrics which are used to calculate the customer intimacy components out of the available data in the provider's information system. Pre-

vious literature provides numerous meaningful indications to perform this activity. For instance, the Industrial and Marketing Purchasing Group (IMP) presented several contributions in which they recognize that relationships are based on organized patterns of interactions (Hakansson & Snehota, 2000, p.75). The identification of these patterns is based on specific characteristics such as the quantity, intensity, and regularity of the interactions. These aspects are, thus, potential customer intimacy metrics for assessing the relationships established between the provider and the customer. In this step of the methodology, the customer intimacy metrics are defined in a generic manner, and the most relevant metrics as well are their respective weights are still unknown.

4. **Calculate the customer intimacy metrics**
   In order to identify which of the generic customer intimacy metrics are most relevant, the next step consists of calculating them at both the organizational and individual levels. This activity corresponds to the data selection task in the knowledge discovery in databases process presented in section 3.2.1. In order to perform this calculation, the software *CI Analytics* has been conceived and implemented in the scope of this thesis. This software retrieves and transforms the available customer data, calculates the customer intimacy metrics, and provides the means to visualize their values. In its current version, this application focuses on data which is available in the customer relationship management system CAS genesisWorld.[1] Further details of the software *CI Analytics* are provided in chapter 6.

5. **Empirically estimate the customer intimacy components**
   Similarly to step 4, this activity also corresponds to the data selection part of the knowledge discovery in databases process. To calibrate the *CI Analytics* model to the individual characteristics of a provider organization, some reference values for each of the customer intimacy components are required. Indeed, this methodology follows the supervised learning approach pre-

---

[1]   See *http://crm.cas-software.com/EN/home.asp* (accessed on 11.11.2011).

sented in section 3.2: the relevance of the customer intimacy metrics which are calculated in step 4 is determined upon some specific target values. The *CI Analytics* methodology proposes to determine these reference values by means of a survey performed with the provider employees. Consequently, a questionnaire enabling the provider employees to estimate the acquired customer intimacy components has been designed. This questionnaire contains multiple items which are presented in section 5.2.4. Widely used Likert-type scales are used to measure the agreement or disagreement of the respondents to each item (Miller & Salkind, 2002, p.330). While further details on the design of the questionnaire are introduced in section 3.1.3, a description of the actual survey performed with employees of CAS Software AG to validate the methodology is provided in chapter 7.

6. **Calibrate the model by applying data-mining techniques**
   The step 6 of the methodology refers to the actual calibration of the *CI Analytics* model. The aim of the calibration is to determine a means to combine the customer intimacy metrics calculated in step 4 in a way that this combination reflects the reference values of the customer intimacy components which have been empirically estimated upon a survey in step 5. In order to perform this task, data-mining techniques which aim at discovering patterns in data sets are applied. Thus, this activity corresponds to the steps pre-processing, transformation, and data-mining of the knowledge discovery in databases process. As explained in section 3.2.2, the machine learning algorithms C4.5, support vector machine, *k*-nearest neighbor, and multi-layer perceptron neural network are considered in the scope of this thesis. Chapter 7 illustrates how this calibration is performed in a real scenario.

7. **Validate and interpret the model**
   The last activity of the *CI Analytics* methodology refers to the validation of the calibrated model and relates to the evaluation and interpretation tasks of the knowledge discovery in databases process. It is necessary to assess the generalization

error of the proposed machine learning models, as well as to confirm that the customer intimacy metrics can be used to assess the customer intimacy components. This evaluation is performed by means of a 10-times 10-fold cross-validation with the performance indicators described in section 3.2.3. Finally, the created machine learning models are interpreted in order to derive some meanings, such as operational and managerial implications out of the proposed calculation of the customer intimacy components values.

The main aspects of the *CI Analytics* model such as the customer intimacy components, the customer intimacy metrics and the customer intimacy data, have been introduced along the description of the *CI Analytics* methodology. The next section summarizes these different components and highlights their relationships.

5.1.2. *CI Analytics* Model

The diagram depicted in figure 5.2 illustrates the *CI Analytics* model. This model consists of three main layers:

1. **Customer Intimacy Layer**
   The first layer, called Customer Intimacy Layer, reflects the results of the breakdown analysis of the concept of customer intimacy in meaningful customer intimacy components based on a literature review. This layer is the outcome of the first step of the *CI Analytics* methodology. This breakdown analysis and the resulting customer intimacy components are detailed in chapter 4.

2. **Customer Intimacy Network**
   The second layer, defined as Customer Intimacy Network, consists of the different customer intimacy metrics which have been designed in order to infer the customer intimacy components. To support the objective to provide an assessment of the customer intimacy components with multiple levels of details, a social network is used for representing the information contained in this layer. As explained in section 3.1.3, in
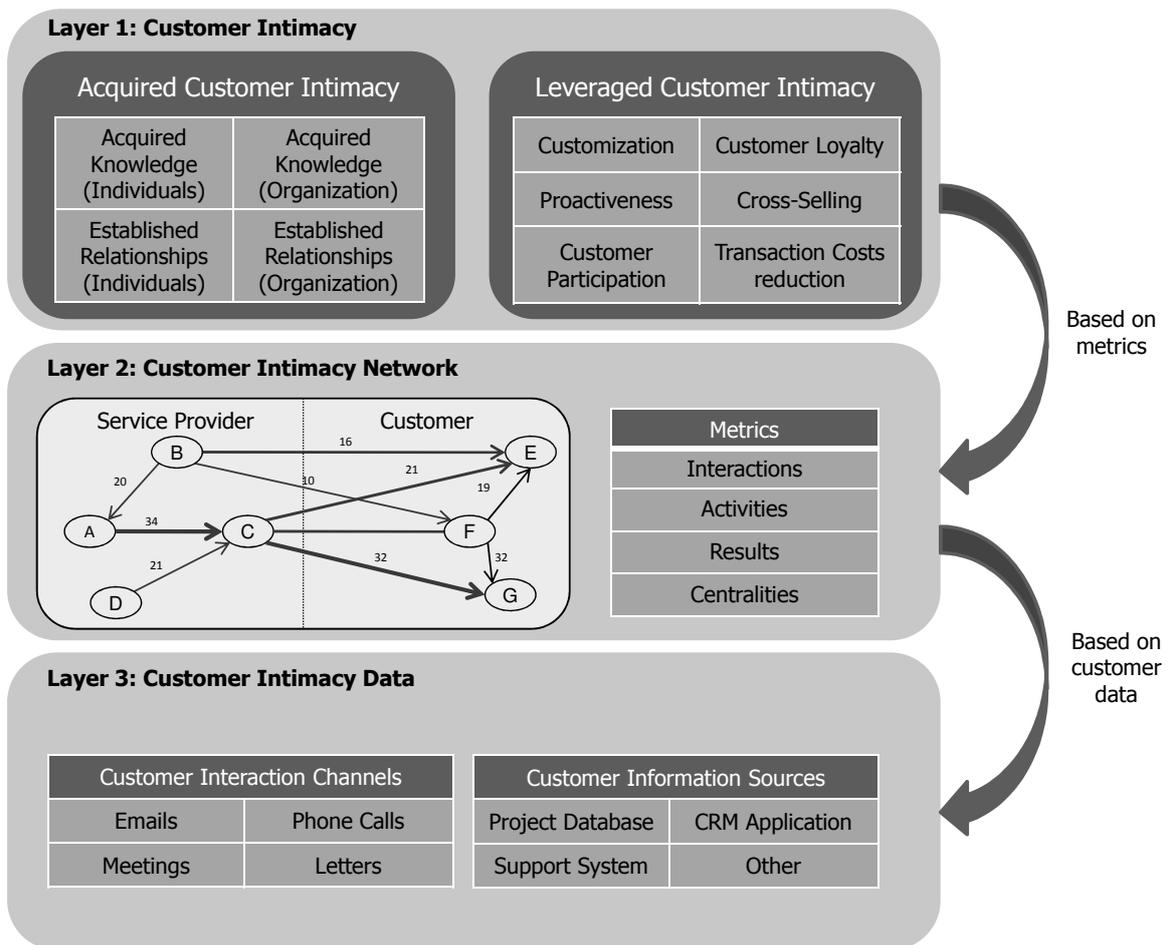
Figure 5.2.: *CI Analytics* Model

this customer intimacy network, the vertices represent the provider and customer employees, and the edges and their respective weights are derived from the different customer intimacy metrics. Moreover, this network representation provides the ability to leverage specific graph based metrics called centrality metrics.[2] These centrality metrics provide a means to aggregate the metrics calculated at the individual level along multiple employees which form a team or a department. Three main types of metrics have been identified and will be further described in sections 5.2 and 5.3:

- *Interaction metrics* focus on the characteristics of the dialog and exchanged informations between the provider and the customer such as interaction regularity, quantity, or intensity.

- *Activity metrics* measure the efforts performed by the provider for the customer, such as the time spent to customize a solution for the customer.

- *Result metrics* focus on the concrete achievements with the different customers, such as sales and projects based metrics.

3. **Customer Intimacy Data**
   The layer Customer Intimacy Data holds the underlying raw data, the "evidence of customer intimacy". Two main types of sources can be distinguished:

   - *Customer interaction channels* consist of the different means used by the provider and the customer in order to exchange information, to dialog, and to jointly perform activities, such as emails, phone calls, letters, and face to face meetings. As it will be explained in section 5.2, much literature confirms the close association between knowledge, relationships, and interactions (Donaldson & O'Toole, 2007; Gummesson, 2008; Hakansson *et al.*, 2009).

---

[2]   Further details on centrality metrics are provided in chapter 3.

- *Customer information sources* contain additional relevant data for the calculation of the customer intimacy components such as the project databases, the customer relationship management system, or the support system storing the requests and issues of the customers.

To summarize the *CI Analytics* model, the customer intimacy components which are specified in the first layer (Customer Intimacy Layer) result from the first step of the *CI Analytics* methodology, which is the breakdown analysis of the concept of customer intimacy. The values of these components are inferred from the metrics proposed in the second layer (Customer Intimacy Network). These metrics are calculated upon existing data which is available in the third layer of the model (Customer Intimacy Data). The identification of this data results from the step 2 of the *CI Analytics* methodology. The step 3 of the *CI Analytics* methodology provides a generic form of the customer intimacy metrics. The remaining steps 4 to 7 of the methodology enable a calibration of the proposed metrics to the specific patterns of each provider.

In the next two sections, the steps 2 and 3 of the *CI Analytics* methodology are detailed for the acquired and for the leveraged customer intimacy: the investigated sources of customer intimacy data as well as the customer intimacy metrics designed to infer the acquired and leveraged customer intimacy components are introduced.

5.2. Assessment of the Acquired Customer Intimacy

It has been established in chapter 4 that the acquired customer intimacy can be broken down into two main components: acquired customer knowledge and established customer relationships. It has also been determined that these two components should be assessed with two levels of analysis: the individual and the organizational levels. The individual level of analysis focuses on the acquired customer intimacy established by provider employees with customer employees

whereas the organizational level focuses on the acquired customer intimacy established between provider employees with customer organizations.

In this section, part 5.2.1 is concerned with the identification of data sources which are relevant for assessing acquired customer knowledge and established customer relationships. Part 5.2.2 and 5.2.3 focus on the actual metrics proposed by this thesis to calculate the values of these two components at the individual and organizational levels. Finally, part 5.2.4 elaborates on the series of Likert-items chosen to empirically assess them.

5.2.1. Using Interactions and Networks to Assess Acquired Customer Intimacy

In the scope of this thesis, the main sources of customer intimacy data for the assessment of the acquired customer intimacy components are the interaction and communication data. The positive correlation between relationships, knowledge, communication, and interaction has already been confirmed in numerous ways in past literature, as explained in the next paragraphs. Moreover, a key aspect of this thesis lies in the application of network theory for the visualization and analysis of this assessment: a graph based representation is used to depict the acquired customer intimacy at the individual level and centrality metrics are calculated on these graphs to aggregate the information up to the organizational level. Both Gummesson and Batt confirm the relevance of this approach: while Gummesson (1995) qualifies relationship marketing as "marketing seen as interactions, relationships and networks," Batt (2004, p.171) states that "interaction is the key construct at the heart of relationship marketing and the network paradigm."

A major stream of research focusing on the analysis of business relationships through the study of interactions is driven by the Industrial and Marketing Purchasing Group (Hakansson & Snehota, 2000; Leek *et al.*, 2001).[3] This group argues that relationships between buy-

---

[3]   See IMP website *http://www.impgroup.org* (accessed on 11.11.2011).

ers and sellers are "built from interaction processes in which technical, social, and economic issues are dealt with" (Hakansson & Snehota, 2000, p.75). In this context, a business interaction is defined as "the process that occurs between companies and which changes and transforms aspects of the resources and activities of the companies involved in it and the companies themselves" (Hakansson *et al.*, 2009, p.27). The underlying theory of the approach followed by the IMP Group is the social / relational exchange theory which perceives relationships as social entities (Donaldson & O'Toole, 2007). From this perspective, the actors are the provider and customer organizations as well as the employees that belong to these organizations. These individuals create multiple interpersonal ties and are involved in multiple dialogs along the development of the interaction process. These various aspects at different levels need to be considered in order to understand the overall relationship (Medlin & Törnroos, 2006). The emphasis is not on the provider side, but on the inter-organizational relationships established between the provider and its customers at both the organizational and individual levels: relationships belong to a social structure and can be analyzed with social network analysis (Granovetter, 1985; Husted, 1994).

Three different research contributions have confirmed the potential of investigating knowledge and relationships through an analysis of social networks. First, Hutt & Walker (2006) have established a means to determine the effectiveness of key account management programs through an analysis of the centrality metrics related to the individual key account managers. Then, focusing on the internal social network established by employees inside a large consulting firm, Wu *et al.* (2009) have determined metrics based on the topology of the social network and correlated them with the success of teams and employees of this company. Finally, Kiss (2007) has developed a model to leverage network data available in customer relationship management systems in order to improve customer classification tasks and to support viral marketing activities.

Another important stream of research on relationships and interactions is conducted in the domain of service marketing by the "Nordic School of Thought," which looks at interactions as a marketing be-

havior (Grönroos, 2000). This perspective emphasizes that relationship marketing is effective if both the provider and the customer consider themselves as being part of a relationship: "a relationship has developed when a customer perceives that a mutual way of thinking exists between customer and supplier or service provider." (Grönroos, 2007, p.36). The development of this relationship in a service context is mainly supported by the interactions and communications that occur between the provider and the customer. More precisely, a two-way communication, a dialog, is essential for the development of the relationship. Ballantyne (2004, p.114) argues that dialog, understood as an "interactive process of learning together", provides the means to obtain business knowledge as well as to develop trust among the business partners.

Under the label of the "Nordic School of Thought", Holmlund (2004) developed a relationship framework in which the relationship is broken down into a flow of interactions in order to analyze its quality. This flow contains three different levels of aggregation, as depicted in figure 5.3:

- *Act* is the smallest element of the interaction flow and represents the "moment of truth" in the relationship. It consists of an actual exchange of information or a joint activity, such as a meeting, a phone call, or an email.

- *Episode* is a series of acts related to the same task, and forming a minor part of the relationship, like a negotiation or a specific part of a project.

- *Sequence* is a series of episodes and represents a major aspect of the relationship like a whole project.

This model is relevant in the context of this thesis as it presents multiple benefits for the analysis of business relationships. According to Holmlund (2004), it allows for a detailed analysis of the associations between different relationship concepts, such as trust, commitment, and value creation. It also provides the ability to compare different relationships based on objective criteria, such as the duration of the overall relationships or the number of sequences. Last, it provides a

good framework for structuring empirical data when quantitative or qualitative data is used.
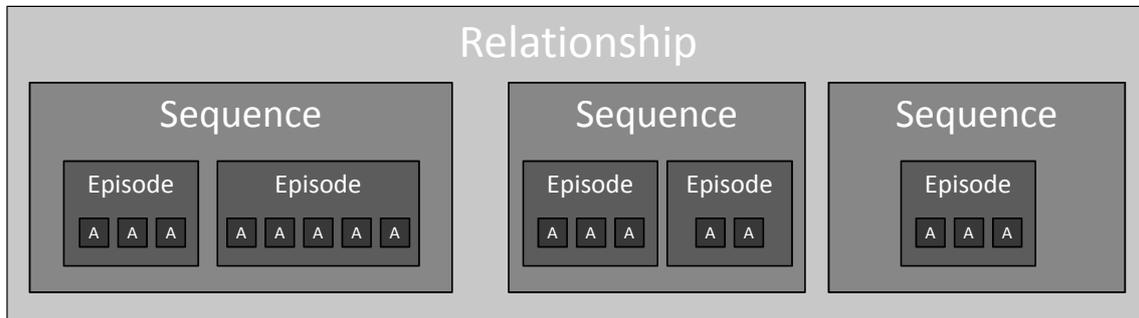
Figure 5.3.: Interaction Levels in a Relationship (Holmlund, 2004)

The next section shows that the metrics created for the assessment of the acquired customer intimacy components are inspired by the previously introduced network perspective and by the decomposition of the relationship in series of sequences, episodes, and acts. The actual interaction and communication data sources which are considered in this thesis are the following:

- Meetings in persons, also called "face to face" meetings, in which provider and customer employees exchange information and knowledge and participate in joint activities;

- Phone calls, which are direct and synchronous communications between providers and customers employees;

- Emails, as an asynchronous type of communication which may involve multiple employees on both the provider and customer sides;

- Letters, as another asynchronous communication channel.

5.2.2. Customer Intimacy Metrics at the Individual Level

The metrics at the individual level are those that assess to which extent a provider employee p knows a customer employee c and has established a relationship with this person. One of the core objectives of this thesis is to provide an automated measurement of these

values based on the following data available in the provider's information system: meetings, phone calls, emails, and letters. Two main challenges reside in this activity:

- First, the confidentiality of the information must be respected: for legal reasons it is not permitted to look into the content of emails or letters, or to analyze the minutes of meetings and phone calls. Therefore, the model proposed in this thesis focuses only on the "existence" of such data but not on the actual content of the data itself. For instance, the information that several meetings between the provider and the customer occurred over the past year can be used. However, the topics of these meetings is not considered.

- The second challenge refers to the inference challenge described in section 5.1. There is no means to know in advance which metrics should be calculated in order to assess the acquired customer intimacy components. For instance, the number of interactions, their duration, as well as the overall duration of the relationship are all potential indicators of these values. The problem is made even more complex as there is no available means to weight the impact of the different interaction and communication channels. For instance, it cannot be argued that a certain amount of face to face meetings is more important than a specific quantity of emails. In order to remedy this issue, the concept of customer interaction time, which is explained in the next paragraph, has been developed.

5.2.2.1. The Concepts of Customer Interaction Time and Weighted Customer Interaction Time

Since it is not possible to estimate ex ante the relative importance of the different customer interaction channels, this thesis proposes to aggregate the overall time spent communicating and interacting with the customer across all different channels in a value which is called customer interaction time (CIT). In order to calculate this CIT value, first, the different acts that belong to the relationships are evaluated in order to calculate their respective contribution to the overall

CIT value. Then, these values are summed along each different interaction channel. Finally, the overall *CIT* value is aggregated as the total customer interaction time across all interaction channels. For instance, if a meeting lasting two hours is followed by two phone calls which last respectively 10 and 20 minutes, the overall customer interaction time is equal to 2 hours and 30 minutes. As opposed to phone calls and meetings, the *CIT* value for emails and letters cannot be directly measured. Still, multiple functions can be defined and calibrated, which take into account for instance the time to write or read them. In this model, as a first approximation, we assume that each email has a constant *CIT* value of d_{email} and each letter has a constant *CIT* value of d_{letter}. Both d_{email} and d_{letter} can be individually calibrated for each provider. In future research, these two constants could be replaced by functions which take multiple parameters into account such as the length of the emails and letters or the roles of the senders and receivers.

An important parameter which determines the quality of the communication and interaction between two individuals is the number of participants to the different interactions in which they are involved. If a provider employee has a one-hour meeting with one single customer employee, he is more likely to obtain knowledge about this person and to establish a relationship with this person than if he meets this person in a larger event with several people involved. In a similar way, if an email is sent by a provider employee to one person, it certainly contains more personalized information than if this email is sent to all employees of the customer organization. Thus, this model takes into account the number of participants to each interaction, and a second calculation of the customer interaction time called *weighted customer interaction time* (*wCIT*) is provided. Further details on the calculation of *CIT* and *wCIT* are presented in the next paragraphs.

The *CIT* and *wCIT* values are not calculated only once for the overall duration of the relationship, but can be evaluated for various time intervals. This feature provides the ability to identify the multiple episodes and sequences that belong to the relationship, as proposed in the relationship framework presented in figure 5.3. Indeed, if no
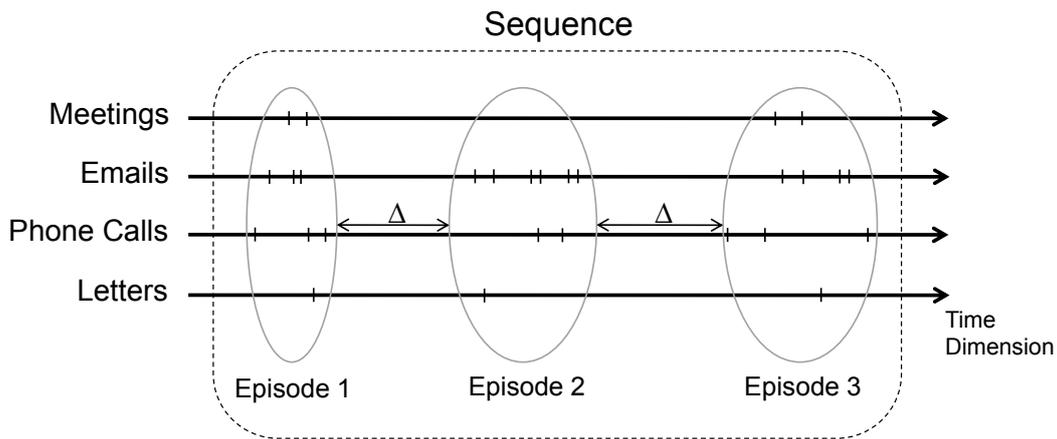
Figure 5.4.: Customer Interaction Time: A Means To Aggregate Customer Interaction Across Multiple Channels

interaction occurred in any of the channels for a duration which is above a certain interaction duration threshold Δ, this model assumes that a new episode has begun. Figure 5.4 illustrates an example of using the customer interaction time to identify the different episodes and their respective compositions. In this example, the first episode of the sequence consists of two meetings, three emails, three phone calls and one letter. Then, no interaction occurs for a time period which is equal to the interaction duration threshold Δ. This fact indicates the beginning of the second episode which consists of six emails, two phone calls and one letters, but no face to face meetings. Finally, the third episode of the sequence starts after the interaction duration threshold has been reached for a second time. It consists of two meetings, four emails, two phone calls, and one letter.

In order to define the customer intimacy metrics within this model, further mathematical formalization is required. In this model, the relationship is analyzed over a time period of duration T. This time period T is divided in multiple contiguous time segments which all have the same duration d. $S = \{s_1, ..., s_i, ..., s_n\}$ represents the set of segments which compose the time period T. Thus, $|S|$, the cardinality of S can be derived from the time period T and segment duration: $|S| = \frac{T}{d}$. A realistic example would be the interaction analysis over the last year, defining the segment size as one month. In this case

$T = 12$ months, $d = 1$ month, and $|S| = 12$. Another option would be to increase the level of detail of the analysis and set the segment size to one week. In this case $|S|$ would be equal to 52.

$CIT_c^p(s_i)$ and $wCIT_c^p(s_i)$ represent the customer interaction time and weighted customer interaction time between the provider employee p and the customer employee c within the time segment s_i. They are calculated as the sum of the customer interaction time (resp. weighted customer interaction time) of all interactions that occurred across the four different channels between these two employees during s_i. If $H = \{meetings, phonecalls, emails, letters\}$ represents the set of interaction channels available to the provider employee p and the customer employee c, then:

$$CIT_p^c(s_i) = \sum_{h \in H} CIT_{p,h}^c(s_i) \tag{5.1}$$

$$wCIT_p^c(s_i) = \sum_{h \in H} wCIT_{p,h}^c(s_i) \tag{5.2}$$

The different components of the equations are calculated as follows:

- $CIT_{p,meetings}^c(s_i)$ and $wCIT_{p,meetings}^c(s_i)$ represent the total time and total weighted time spent in meetings in which the provider p and c participated. If $K_{p,meetings}^c(s_i)$ symbolizes the set of meetings within the time segment s_i in which both the provider employee p and the customer employee c participated, d_j the duration of the meeting j, and n_j the number of participants to the meeting j, without counting the employee p, then:

$$CIT_{p,meetings}^c(s_i) = \sum_{j \in K_{p,meetings}^c(s_i)} d_j, \tag{5.3}$$

$$wCIT_{p,meetings}^c(s_i) = \sum_{j \in K_{p,meetings}^c(s_i)} \frac{d_j}{n_j} \tag{5.4}$$

- $CIT^c_{p,phonecalls}(s_i)$ and $wCIT^c_{p,phonecalls}(s_i)$ represent to the total time and total weighted time spent in phone calls in which the provider p and c participated. If $K^c_{p,phonecalls}(s_i)$ symbolizes the set of phone calls within the time segment s_i in which both the provider employee p and the customer employee c participated, d_j the duration of the phone call j, and n_j the number of participants to the phone call j, without counting the employee p, then:

$$CIT^c_{p,phonecalls}(s_i) = \sum_{j \in K^c_{p,phonecalls}(s_i)} d_j \tag{5.5}$$

$$wCIT^c_{p,phonecalls}(s_i) = \sum_{j \in K^c_{p,phonecalls}(s_i)} \frac{d_j}{n_j} \tag{5.6}$$

- $CIT^c_{p,emails}(s_i)$ and $wCIT^c_{p,emails}(s_i)$ represent the importance of email communication. If $K^c_{p,emails}(s_i)$ symbolizes the set of emails exchanged between p and c within the time segment s_i, and n_j the number of recipients of the email j, then:

$$CIT^c_{p,emails}(s_i) = |K^c_{p,emails}(s_i)|.d_{email} \tag{5.7}$$

$$wCIT^c_{p,emails}(s_i) = \sum_{j \in K^c_{p,emails}(s_i)} \frac{d_{email}}{n_j} \tag{5.8}$$

- $CIT^c_{p,letters}(s_i)$ and $wCIT^c_{p,letters}(s_i)$ represent the importance of mail communication. If $K^c_{p,letters}(s_i)$ symbolizes the set of letters exchanged between p and c within the time segment s_i, and n_j the number of recipients of the letter j, then:

$$CIT^c_{p,letters}(s_i) = |K^c_{p,letters}(s_i)|.d_{letter} \tag{5.9}$$

$$wCIT^c_{p,letters}(s_i) = \sum_{j \in K^c_{p,letters}(s_i)} \frac{d_{letter}}{n_j} \tag{5.10}$$

In order to identify the interaction episodes between the provider employee p and the customer employee c, the set of episodes within the time period T is denoted EP_p^c. The principle for the identification of the different episodes is as follows: the previously introduced interaction duration threshold Δ is set as proportional to the segment duration d: $\Delta = \lambda \times d, \lambda \in \mathbb{N}$. If no interaction occurs within contiguous segments whose total duration is equal to Δ, then the next interaction indicates the beginning of a new episode. In this model, Δ is equal to the segment duration d ($\lambda = 1$). Thus, if no interaction occurs within one time segment, the next interaction indicates the beginning of a new episode. The length of the episode is, consequently, determined by the number of contiguous segments in which some interaction occurred.

A restriction in this model is that, so far, the actual CIT and $wCIT$ values are not considered for the identification of the episodes. Even a small interaction, like one email, is sufficient in order to have the segment it belongs to being considered as part of an episode. Therefore, two additional threshold values, the interaction quantity threshold b and the weighted interaction quantity threshold wb have been created. Using these parameters, a segment s_i is ignored in the identification of the episode if the corresponding customer interaction time CIT and weighted customer interaction time $wCIT$ are below their respective threshold b and wb: $CIT_p^c(s_i) < b$ and $wCIT_p^c(s_i) < wb$

Figure 5.5 illustrates the identification of the different episodes upon the analysis of the segments. In this example, b and wb are both set to zero and Δ is set to the segment duration d. The first episode e_1 is mapped to the first segment s_1. Then, no interaction occurs with the segment s_2 and, therefore, a new episode starts with the segment s_3. Both s_3 and s_4 contain multiple interactions. Therefore, the second episode is spread over these two segments. Since the segment $s5$ and s_6 do not contain any interaction, the third episode starts with the segment s_7. Consequently, using this approach, three episodes can be identified over this time frame consisting of seven segments. Importantly, the length of the analyzed time frame T as well as the segment size d have a significant role in the identification
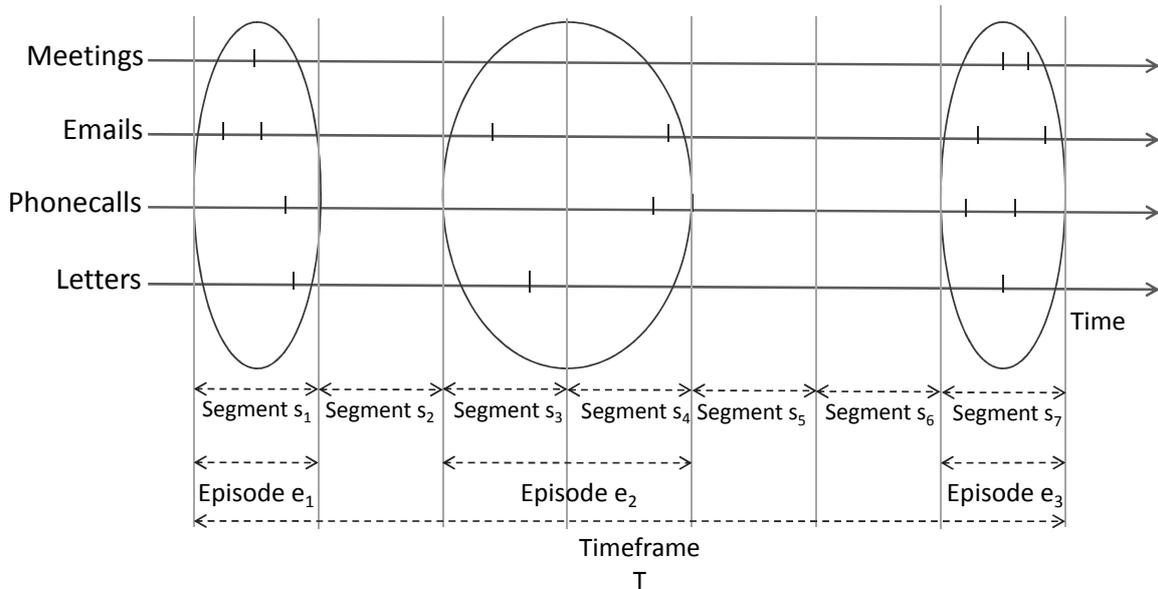
Figure 5.5.: Segmentation of the Relationship to Identify Episodes Across Multiple Channels

of the episodes and, thus, have to be carefully chosen at the calibration time.

5.2.2.2. Acquired Customer Intimacy Assessment at the Individual Level

The metrics conceived in this thesis to assess the acquired customer intimacy are based upon four interaction characteristics which have already been identified in past literature: quantity, intensity, regularity, and mode of interaction. These characteristics will be further developed in the next paragraphs. Based on the previously introduced concepts of segment, episode, customer interaction time, and weighted customer interaction time, eight metrics which reflect these four interaction characteristics have been created.

Since some of these metrics consider only the most relevant segments within the time frame T, we define S_p^c as a subset of the previously defined set S of segments which exclusively includes segments for which CIT_p^c and $wCIT_p^c$ are above the previously mentioned interac-

tion quantity thresholds b and wb:

$$S_p^c = \{s_i \in S \mid CIT_p^c(s_i) > b \wedge wCIT_p^c(s_i) > wb\} \tag{5.11}$$

1. **Quantity: volume and weighted volume**
   The first metrics, called volume and weighted volume, relate to the quantity of interactions between the provider employee p and the customer employee c. Interaction quantity has already been investigated as a potential indicator of relationships in past literature (Nezlek, 2003): it is more likely that p has established a relationship with c and has acquired some knowledge about c if p and c have a high volume of interaction rather than a low interaction quantity. Volume (resp. weighted volume) is calculated as the customer interaction time (resp. weighted customer interaction time) between these two individuals along the time frame T. The calculation of these two metrics is based on the following equations:

$$Volume_p^c = \sum_{s_i \in S} CIT_p^c(s_i) \tag{5.12}$$

$$wVolume_p^c = \sum_{s_i \in S} wCIT_p^c(s_i) \tag{5.13}$$

2. **Intensity and weighted intensity**
   The second metric refers to the intensity of the interaction. A certain volume of interaction can be reached upon either multiple small and sporadic low-intensity acts, or with a limited number of acts with a higher intensity. The influence of the interaction intensity on knowledge flow and relationships has also been studied in multiple ways in a business context. Noorderhaven & Harzing (2009, p.2) identify in their research that "intensive social interaction provides opportunities for social construction of knowledge in a learning dialogue". Similarly, Bennett & Robson (1999) argue that intense interactions support the exchange of information and is a means to overcome the challenge of knowledge and information asymmetry. Finally, Hakansson *et al.* (2009, p.81) confirm that interaction in-

tensity influences the effects of the interaction on the involved resources from both the provider and the supplier. The metrics intensity ($Intensity_p^c$) and weighted intensity ($wIntensity_p^c$) are calculated as the average customer interaction time (resp. weighted customer interaction time) calculated over the segments which belong to S_p^c:

$$Intensity_p^c = \frac{\sum_{s_i \in S_p^c} CIT_p^c(s_i)}{|S_p^c|} \tag{5.14}$$

$$wIntensity_p^c = \frac{\sum_{s_i \in S_p^c} wCIT_p^c(s_i)}{|S_p^c|} \tag{5.15}$$

3. **Regularity: frequency, duration and number of episodes**[4]
   The third interaction characteristic considered in this model concerns the regularity of the interactions. Regular communication and interaction have been recognized as a key aspect of successful relationship marketing (Berry, 1995). Kong & Mayo (1993) confirm the importance of the regularity dimension as they argue that 'successful business-to-business relationships are based on regular, constructive and innovative interaction." Focusing on communication effectiveness in professional services, Sharma & Patterson (1999, p.163) consider that "regular communications can help develop a sense of closeness and ease in the relationship, and be instrumental in building emotional and social bonds." Therefore, three metrics have been conceived in order to assess the regularity of the interactions: frequency, duration, and number of episodes.

   - Frequency refers to the proportion of segments in which some interactions between the provider employee p and the customer employee c happened within the time period T. If p and c communicated and interacted in multiple different segments, they are more likely to have a regular

---

[4] The weighted version of these metrics are not necessary as their values would be equal to their corresponding "non weighted" versions.

interaction than if they only interacted within one or two segments. Frequency ($Frequency_p^c$) is calculated as the percentage of segments that belong to S_p^c to the total number of segments within the time frame T:

$$Frequency_p^c = \frac{|S_p^c|}{|S|} \tag{5.16}$$

- Duration indicates to which extent the interactions between p and c span over the time frame T. A certain frequency value indicates the number of segments in which some interactions occurred, but it does not specify whether these segments are concentrated in a specific part of T, such as the beginning or the end of T, or if they are uniformly distributed over T. This aspect, however, is significant in the interpretation of the interactions: if the segments are contiguous to each other, it means that p and c had regular interactions over a limited period of time only. On the opposite, if the interactions happened in the first and last segments of the time period T, p and c certainly had more regular interactions. In order to calculate the metric duration, it is necessary to identify the index of the first and last segments of time period T which contain relevant interactions. Considering the set of segments S and its subset S_p^c both chronologically ordered, we define f as the index in S of the first item in S_p^c and l as the index in S of the last item in S_p^c. For instance, if the first and last relevant interactions occurred respectively in the third and seventh segments, then f=3 and l=7. Using these values, the metric duration ($Duration_p^c$) is calculated as follows:

$$Duration_p^c = \frac{l - f + 1}{|S|} \tag{5.17}$$

- Number of episodes is the last conceived metric as indicator of the interaction regularity. It is derived from the rela-

tionship framework presented in figure 5.3. Considering a certain frequency value, a high number of episodes would indicate several interruptions in the relationship while a low number of episodes denotes a continuity in the interaction as several segments would be contiguous to each other. The metric number of episodes ($NumberEpisodes_p^c$) is indicated by the cardinality of the previously introduced set of episodes EP:

$$NumberEpisodes_p^c = |EP_p^c| \qquad (5.18)$$

4. **Mode of interaction**
   The metric mode of interaction ($Mode_p^c$) indicates the proportion of face to face meetings in the overall interaction between the provider employee p and the customer employee c. This metric is derived from the finding that meetings in person have a higher significance in the construction of the relationship and in the exchange of knowledge than the other channels of interaction such as phone calls and emails. In an analysis of 35 sales and services virtual teams, Kirkman *et al.* (2004) identify that the number of face to face meetings is a moderating factor between the teams' empowerment and their performance. Noorderhaven & Harzing (2009, p.2) confirm that "face-to-face social interactions form a communication channel particularly conducive to the transfer of tacit, non-codified knowledge." Mode of interaction is calculated with the following equation:

$$Mode_p^c = \frac{\sum_{s_i \in S} CIT_{p,meetings}^c(s_i)}{\sum_{s_i \in S} CIT_p^c(s_i)} \qquad (5.19)$$

5.2.2.3. A Graph-Based Representation of the Customer Intimacy Metrics

The customer intimacy metrics defined in the previous paragraph are the indicators of some specific interaction patterns between a provider employee and a customer employee. Taking the broader

perspective of the "many-to-many" interactions occurring between all employees of the provider P and of the customer C (Gummesson, 2008), it is possible to use these metrics in order to design multiple graph-based representations of the social network established between the two companies. In these graphs, the vertices are the provider and customer employees, the edges indicate some interactions among these employees, and the weights of the edges are calculated as a function of the previously defined customer intimacy metrics.

The set of graphs that can be inferred out of the customer intimacy metrics is defined as $G = \{G_1, G_k, ..., G_n\}$. G_k is the graph that uses the "weighting" function ω_k, as explained in section 3.1.1, to calculate the weights of the edges. More formally, the graph G_k is defined as $G_k = (V, E_k)$. V is the set of vertices of the graph G_k and consists of the employees involved in the interaction between P and C. V is composed of the two subsets V_C and V_P where V_P represents the employees of the provider organization and V_C the employees of the customer organization: $V = V_P \cup V_C$. If $e_{p,c}$ represents the edge between the provider employee p and the customer employee c, and $w_{p,c}$ the weight of the edge $e_{p,c}$, then E_k represents the set of edges in the graph G_k which link the provider and customer employees and whose weights are calculated with the function ω_k:

$$E_k = \{e_{p,c}; w_{p,c} = \omega_k(e_{p,c}) \,|\, \forall p \in V_P; \forall c \in V_C; \omega_k : E_k \to \mathbb{R}^+\} \quad (5.20)$$

The customer intimacy metrics presented in the previous paragraphs represent some standard functions for calculating the weights of the edges. Indeed, the weights could be defined by the volume or by the intensity of the interactions between the provider employee p and the customer employee c. In these cases, the weighting function would be respectively:

$$\omega_{Vol}(e_{p,c}) = Volume_p^c \quad (5.21)$$

$$\omega_{Int}(e_{p,c}) = Intensity_p^c \quad (5.22)$$

Figure 5.6 illustrates the representations of the social network formed by provider and customer employees using the metrics volume and intensity as weighting functions. While the volume of interaction is
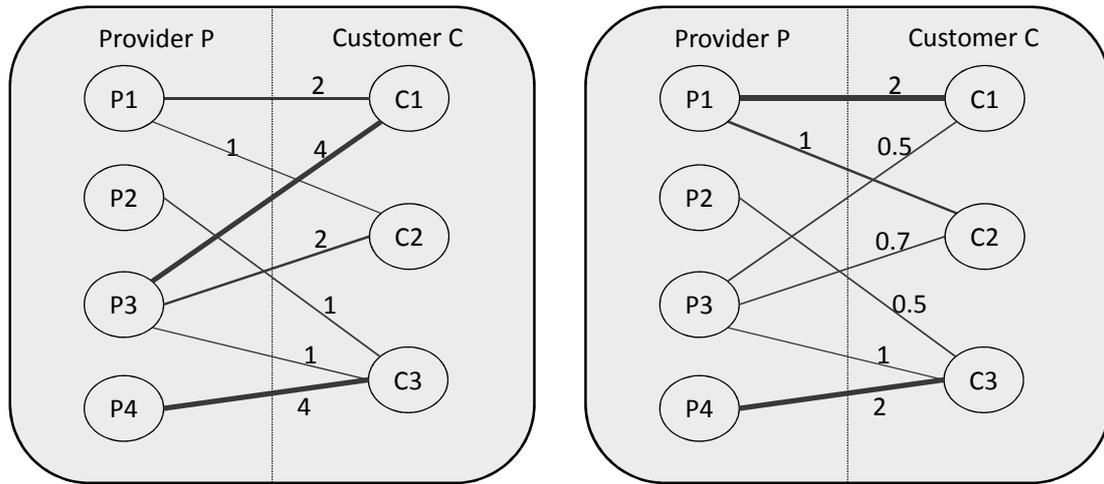
(a) weighting function based on volume (b) weighting function based on intensity

Figure 5.6.: Two Different Graph Representations of the Social Network Formed by the Provider and Customer Employees

used to calculate the weights of the edges in graph 5.6(a), the intensity is used in graph 5.6(b). The provider and customer employees involved in the interaction are: $V_P = \{P1, P2, P3, P4\}$ and $V_C = \{C1, C2, C3\}$. It can be observed that the weights of the edges are significantly different on both graphs. For instance, in graph 5.6(a) the edges $e_{P3,C1}$ and $e_{P4,C3}$ both indicates higher volumes of interaction than the other edges. However, in graph 5.6(b), it is shown that the edge $e_{P4,C3}$ has a high value of intensity but the edge $e_{P3,C1}$ has a low value of intensity. This example illustrates that the chosen weighting function significantly impact the resulting representation of the social network formed by the provider and customer employees.

This thesis aims at leveraging the conceived customer intimacy metrics in order to infer the values of the customer intimacy components, which are at the individual level, the acquired knowledge of, and established relationships with, customer employees. The objective from a social network analysis perspective is, thus, to determine the two weighting functions $\omega_{Knowledge}$ and $\omega_{Relationship}$, so that the weights of the edges in the graph representation indicate respectively

the acquired knowledge of, and established relationships with, customer employees. In order to identify these weighting functions, the steps 4 to 7 of the *CI Analytics* methodology are performed. These steps are detailed in chapter 7.

The next section elaborates on the metrics conceived to assess the acquired customer intimacy at the organizational level.

5.2.3. Customer Intimacy Metrics at the Organizational Level

Focusing on the acquired customer intimacy part of the *CI Analytics* model presented in figure 5.2, the metrics at the individual level indicate to which extent employees of the provider organization have established relationships with, and acquired knowledge of, customer employees. Similarly, as explained in chapter 4, the metrics at the organizational level indicate the relationship that a provider employee has established with a specific customer organization as well as the amount of knowledge he has acquired on this organization.

Two different types of metrics are proposed by this thesis to assess the acquired customer intimacy at the organizational level:

- first, the concepts of customer interaction time and weighted customer interaction time are adapted so that an assessment of these values at the organizational level is applicable. Thus, the metrics at the individual level presented in section 5.2.2 can also be calculated at the organizational level.

- second, further metrics which leverage the characteristics of the previously introduced social network established between the provider and customer organizations are defined. These metrics are the degree centrality, the normalized degree centrality, and the normalized closeness centrality.

5.2.3.1. The Concept of Customer Interaction Time at the Organizational Level

The previously introduced customer interaction time $CIT_p^{c_1}(s_i)$ and weighted customer interaction time $CIT_p^{c_1}(s_i)$ are calculated using

some specific aggregation functions of the interactions occurring between the provider employee p and the customer employee c_1 across all interaction channels within the time segment s_i. During the segment s_i, the provider employee p may not only have interacted with the customer employee c_1, but also with the employees c_2 and c_3 from the same customer organization C. Consequently, it is possible to calculate in the same manner the customer interaction time and weighted customer interaction time between p and c_2 and between p and c_3, leading to the values $CIT_p^{c_2}(s_i)$, $wCIT_p^{c_2}(s_i)$, $CIT_p^{c_3}(s_i)$ and $wCIT_p^{c_3}(s_i)$. By aggregating these multiple values, it is possible to obtain the customer interaction time and weighted interaction time between the provider employee p and the customer organization C or any of its subsets such as the team C_1 formed by the employees c_1, c_2, and c_3.

More formally, the set of employees of the customer organization C is defined as V_C. If C_x represents the subset x of the organization C, such as a team, a department, or a business unit, V_{C_x} represents the set of employees which belong to C_x. Thus V_{C_x} is either included in V_C or equal to V_C: $V_{C_x} \subseteq V_C$. It is then possible to calculate the CIT and $wCIT$ values between the provider employee p and C_x with the following equations:

$$CIT_p^{C_x}(s_i) = \sum_{c \in V_{C_x}} CIT_p^c(s_i) \qquad (5.23)$$

$$wCIT_p^{C_x}(s_i) = \sum_{c \in V_{C_x}} wCIT_p^c(s_i) \qquad (5.24)$$

It is also possible to calculate the customer interaction time and weighted customer interaction time for any specific interaction channel at the organizational level. As defined previously, if $H = \{meetings, phonecalls, emails, letters\}$ represents the set of interaction channels available to the provider employee p for interacting with employees of the customer organization C, then:

$$CIT_{p,h}^{C_x}(s_i) = \sum_{c \in V_{C_x}} CIT_{p,h}^c(s_i) | \forall h \in H \qquad (5.25)$$

$$wCIT^{C_x}_{p,h}(s_i) = \sum_{c \in V_{C_x}} wCIT^c_{p,h}(s_i) | \forall h \in H \qquad (5.26)$$

5.2.3.2. Organizational Metrics Based On Customer Interaction Time

Using the concepts of customer interaction time CIT and weighted customer interaction time $wCIT$ applied at the organizational level, it is possible to calculate the metrics presented in section 5.2.2 for any organization C or subset of this organization C_x, such as a team, a department or a business unit. Since these metrics are described and motivated in detail in section 5.2.2, this section mainly outlines the equation defined in order to adapt them to the organizational level. In order to provide the ability to ignore the time segments in which a certain level of interaction has not been reached for the calculation of the customer intimacy metrics, the interaction quantity thresholds B and wB are defined at the organizational level. $S^{C_x}_p$ represents the subset of time segments within the time period T for which the overall customer interaction time for all employees that belong to the organization C_x (resp. weighted customer interaction time) is above the threshold B (resp. wB):

$$S^{C_x}_p = \{s_i \in S | CIT^{C_x}_p(s_i) > B \land CIT^{C_x}_p(s_i) > wB\} \qquad (5.27)$$

The customer intimacy metrics at the organizational level which leverage the concepts of customer interaction time and weighted customer interaction time are based on the four following interaction characteristics: quantity, intensity, regularity, and mode of interaction.

1. **Quantity: volume and weighted volume**
   The metrics volume ($Volume^{C_x}_p$) and weighted volume ($wVolume^{C_x}_p$) are indicators of the quantity of interaction that occurred between the provider employee p and the subset C_x of the organization C. These values are calculated as the sum of the customer interaction time (resp. weighted customer interaction time) for all customer employees that belong to the organization subset C_x along all time segments s_i in the time period T:

$$Volume_p^{C_x} = \sum_{s_i \in S} CIT_p^{C_x}(s_i) \tag{5.28}$$

$$wVolume_p^{C_x} = \sum_{s_i \in S} wCIT_p^{C_x}(s_i) \tag{5.29}$$

2. **Intensity and weighted intensity**
   The metrics intensity ($Intensity_p^{C_x}$) and weighted intensity ($wIntensity_p^{C_x}$) indicate whether the relationship with the customer employees that belong to C_x is based on multiple small interactions or on fewer acts which have a longer duration. Their calculation, thus, is based on the average customer interaction time (resp. weighted customer interaction time) along the relevant time segments which belong to $S_p^{C_x}$.

$$Intensity_p^{C_x} = \frac{\sum_{s_i \in S_p^{C_x}} CIT_p^{C_x}(s_i)}{|S_p^{C_x}|} \tag{5.30}$$

$$wIntensity_p^{C_x} = \frac{\sum_{s_i \in S_p^{C_x}} wCIT_p^{C_x}(s_i)}{|S_p^{C_x}|} \tag{5.31}$$

3. **Regularity: frequency, duration, number of episodes**
   The regularity of the interaction indicates whether the interactions with the customer employees are evenly spread along the different time segments or if they are concentrated in some specific segments, for instance at the beginning of at the end of the time period T. In order to assess the interaction regularity, three metrics have been defined: frequency, duration, and number of episodes:

   - Frequency ($Frequency_p^{C_x}$) represents the proportion of time segments in which relevant interactions occurred between the provider employee p and employees of the customer that belong to C_x within the time period T. It is calculated as follows:

$$Frequency_p^{C_x} = \frac{|S_p^c|}{|S|} \tag{5.32}$$

- Duration ($Duration_p^{C_x}$) describes to which extent the interactions between the employee p and employees of the customer organization C_x span over the time period T. If f and l represent respectively the index of the first and last segments which contain some relevant interactions between p and C_x, then the metric duration is calculated as follows:

$$Duration_p^{C_x} = \frac{l - f + 1}{|S|} \tag{5.33}$$

- Number of episodes ($NumberEpisodes_p^{C_x}$) relates to the continuity of the relationship, as a small number of episodes indicates that the segments in which some relevant interaction occurred are contiguous to each other, while a higher number of episodes indicate some interruptions between the different interactions. If $EP_p^{C_x}$ represents the set of episode between the provider employee p and the customer organization C_x, the metric number of episodes is calculated as the cardinality of $EP_p^{C_x}$:

$$NumberEpisodes_p^{C_x} = |EP_p^{C_x}| \tag{5.34}$$

4. **Mode of interaction**

   The metric mode of interaction $Mode_p^{C_x}$ refers to the proportion of time spent with meetings in person between the provider employee p and the employees of the customer organization C_x on the overall interaction time. It is calculated with the following equation:

$$Mode_p^{C_x} = \frac{\sum_{s_i \in S} CIT_{p,meetings}^{C_x}(s_i)}{\sum_{s_i \in S} CIT_p^{C_x}(s_i)} \tag{5.35}$$

5.2.3.3. Organizational Metrics Based On Network Theory

In addition to the eight previously defined metrics which are based on the concepts of customer interaction time and weighted customer

interaction time, it is also possible to leverage the graph based representations presented in section 5.2.2. The main advantage of these additional metrics is that they take into account not only the customer employees with whom some interaction occurred, but also all the remaining potential customer employees with whom no relationship so far have been established. These are, thus, further indicators of the integration of the provider organization in the customer organization. For each graph representation G_k of the social network established between the provider and customer organizations P and C, it is possible to calculate the following metrics:

- **Number of contacts (degree centrality)**
  The metric number of contacts ($NumberContacts_k(p, C_x)$) is calculated as the degree centrality of the provider employee p with the customer organization C_x on the graph G_k. This metric is presented in section 3.1.2. Using the metric number of contact, it is possible to determine the number of customer employees with whom a provider employee interacted over a specific time period. This metric, thus, provides the ability to compare the relationship established by different provider employees with the customer organization C_x.

- **Normalized degree centrality**
  The normalized degree centrality of a node i is presented in section 3.1.2. It is defined as the number of edges incident to i divided by the maximum number of potential nodes adjacent to i. Within this model, the normalized degree centrality of the provider employee p with the organization C_x on the graph G_k is denoted $C'^k_{degree}(p, C_x)$. It indicates the proportion of individuals in the organization C with whom the employee p has established some relationships. The normalized degree centrality enables the comparison of the relationship established by a specific provider employee p with multiple customer organizations which all have a different number of employees.

- **Normalized closeness centrality**
  The normalized closeness centrality is defined in section 3.1.2. It has been established in past literature that closeness central-

ity is an indicator of the effectiveness of the ability to communicate and convey information within a defined network (Freeman, 1979; Beauchamp, 1965). Within this model, the closeness centrality ($C'^k_{close}(p, C_x)$) of the employee p with the organization C_x on the graph G_k complements the metrics intensity and volume and indicates the proximity of the employee p with the organization C_x.

In summary, eight metrics have been conceived to assess the acquired customer intimacy at the individual level and 11 metrics have been conceived to assess it at the organizational level. These metrics are listed in table 5.1.

The next section elaborates on the series of Likert-items proposed by this thesis in order to empirically assess the acquired customer knowledge and established customer relationships.

5.2.4. Empirical Assessment of the Acquired Customer Intimacy

As explained in section 5.1.1, the fifth step of the *CI Analytics* methodology requires to perform an empirical assessment of the customer intimacy components acquired customer knowledge and established customer relationships in order to complete the calibration of the customer intimacy metrics. To perform this assessment at both the individual and organizational levels, a series of items assessed on Likert-type scales has been conceived. Further information on Likert-type scales is presented in section 3.1.3 and an illustrative questionnaire is proposed in appendix A.

- **Acquired Customer Knowledge**
  In order to ensure the relevance and validity of the items used in this thesis to empirically assess acquired customer knowledge, these items are derived from the well recognized scales created by Gwinner *et al.* (2005), Jayachandran *et al.* (2004), and Joshi & Sharma (2004). Gwinner *et al.* (2005) assessed customer knowledge with two items in order to determine the influence

Table 5.1.: Customer Intimacy Metrics at the Individual and Organizational Levels

Customer Intimacy Metric	Individual Level	Organizational Level
Interaction Quantity		
Volume	✓	✓
Weighted Volume	✓	✓
Interaction Intensity		
Intensity	✓	✓
Weighted Intensity	✓	✓
Interaction Regularity		
Frequency	✓	✓
Duration	✓	✓
Number of Episodes	✓	✓
Mode of Interaction		
Mode	✓	✓
Network Centrality		
Number of Contacts (Degree Centrality)		✓
Normalized Degree Centrality		✓
Normalized Closeness Centrality		✓

of employee adaptiveness on service customization. Jayachandran *et al.* (2004) empirically estimated the customer knowledge process on a six-item scale in order to determine its influence on customer response capability. Finally, Joshi & Sharma (2004) created a five-item scale to assess customer knowledge development and to evaluate its impact on new product performance.

At the individual level, the following two assertions have been derived from these scales to assess acquired customer knowledge:

1. My knowledge of [CustomerEmployeeName]'s needs is thorough.

2. I learned a lot about [CustomerEmployeeName]'s preferences in the period I worked with him/her.

At the organizational level, the following three items have been used:

1. My knowledge of [CompanyName]'s needs is thorough.

2. I learned a lot about [CompanyName]'s preferences in the period I worked with it.

3. I know the customer [CompanyName] very well.

- **Established Customer Relationships**
  The items used to empirically evaluate the relationships established with customers are inspired by the scales for assessing relationship in a B2B context proposed by Crosby *et al.* (1990), De Wulf *et al.* (2001), and Wuyts & Geyskens (2005). Crosby *et al.* (1990) estimated relationship quality at the employee level in order to determine its influence on services selling. De Wulf *et al.* (2001) considered relationship quality as an outcome and analyzed which factors have some influence on it such as interpersonal communication and preferential treatment. Last, Wuyts & Geyskens (2005) investigated the formation of buyer-supplier relationship and the influence of relationship quality on partner selection.

  Two items have been derived from these scales in order to assess the established customer relationships at the individual level:

  1. I have a high-quality relationship with [CustomerEmployeeName].

  2. I have a very collaborative relationship with [CustomerEmployeeName].

At the organizational level, the following three items have been used:

1. As an employee, I have a high-quality relationship with [CompanyName].

2. As an employee, I have a very collaborative relationship with [CompanyName].

3. I am satisfied with the relationship I have with [Company-Name].

The next section of this chapter elaborates on the metrics defined to assess the leveraged customer intimacy components presented in section 4.4.

5.3. Assessment of the Leveraged Customer Intimacy

Section 4.4 elaborates on six leveraged customer intimacy components which reflect the actual benefits and competitive advantages derived from the acquired customer intimacy. These six components are: customization, customer loyalty, proactiveness, cross-selling, customer participation, and transaction costs reduction. Following the approach of this thesis to assess customer intimacy upon customer related data available in the provider's information system, the six parts of this section define multiple metrics in order to measure these leveraged customer intimacy components.

While the metrics proposed by this thesis to assess the acquired customer intimacy cover both the individual and the organizational levels, the leveraged customer intimacy metrics solely focus on the organizational level for the following reasons. First, an assessment of the leveraged customer intimacy at the individual level for each customer employee would not be useful in a B2B context as the value proposition of the provider targets the customer organization rather than the different customer employees. Second, the data which is available for calculating the leveraged customer intimacy metrics

such as project records and sales results is specified at the organizational level only and not at the individual level. Thus, no data is available to calculate the leveraged customer intimacy metrics at the individual level. Table 5.2 presents the eight metrics proposed by this thesis to assess the leveraged customer intimacy components.

Table 5.2.: Customer Intimacy Metrics for the Leveraged Customer Intimacy

Customization	**Loyalty**
Customization Revenue Ratio	*Customer Purchase Frequency Ratio*
Proactiveness	**Cross-Selling**
Proactiveness Ratio	*Cross-Selling Revenue Ratio*
	Cross-Selling Diversity Ratio
Customer Participation	**Transaction Costs Reduction**
Customer Participation Quantity	*Transaction Effectiveness Ratio*
Customer Participation Ratio	

5.3.1. Customization

As explained in section 4.4.1, mass customization and customerization are different from customization in the context of customer intimacy. Therefore, proposed approaches for measuring mass customization and customerization such as those from Tu *et al.* (2007) or Kumar (2005) are not suited in the scope of this thesis. In line with the *service customization through employee adaptiveness* model, this thesis focuses on customization achieved by provider employees, such as the completion of individual projects and the adaptation of existing solutions. It is possible to determine the revenues derived by such projects and to compare them to those derived from standard products and services such as software licenses in an IT context. This thesis, thus, proposes to measure the degree of customization with

a customer as the ratio between the revenues generated from individually performed services for this customer and the total revenues generated with this customer.

The overall duration T of the relationship between the provider and the customer is divided in multiple time segments: $T = \{1, ..., i, .., n\}$. $R_{standard}(i)$ represents the revenues generated by standard products and services in the time segment i and $R_{custom}(i)$ the revenues generated by customized offering, such as project revenues in the same time segment i. The *customization revenue ratio* can subsequently be calculated as follows for the time segment i:

$$Customization\ Revenue\ Ratio(i) = \frac{R_{custom}(i)}{R_{standard}(i) + R_{custom}(i)} \qquad (5.36)$$

5.3.2. Customer Loyalty

Different approaches have been proposed in order to measure the degree of loyalty of a customer. Dick & Basu (1994) propose to assess customer loyalty by means of a two-dimensional matrix. The first dimension of this matrix – "repeat patronage" – indicates the intention of the customer to repurchase products or services from the same provider or brand. The second dimension of the matrix – "relative attitude" – refers to the behavior of the customer with regard to the provider organization and to the products and services he purchased from this provider. Dick & Basu (1994) argue that loyalty is high when repeat patronage is high and when the relative attitude towards the provider is strongly favorable. Building upon this research, Bennett & Bove (2001) propose to assess the recommendations and referrals performed by the customer to other organizations as well as to consider the customer's repurchasing behavior. In a similar way, Reichheld (2003, p.5) developed the metric "net promoter score" and argues that this one dimensional construct is strongly correlated with high loyalty. In order to assess the net promoter score, customers are asked to answer the following question on a scale from 1 to 10: "How likely is it that you would recommend our company to a friend or colleague?" Customers answering with a

value comprised between 9 and 10 are those which are very loyal to the provider.

Consequently, this thesis proposes to apply these previously defined approaches in order to assess the degree of customer loyalty. Three different means are considered:

- Count the number of recommendations performed by customers and more specifically those leading to additional revenues with new customers. This information can be obtained if customers inform the provider about the recommendations they performed, or if prospective customers indicate that they contacted the provider on the recommendation of another company. Both solutions are easily implementable and can be fostered by a recommendation reward mechanism.

- Apply the net promoter score approach and survey customers with regards to their intention to recommend the provider.

- Assess the behavior of the customer with regard to the frequency of his purchases over the past years. This information can easily be derived from the sales results available in project databases and in the CRM system. This solution is favored as it corresponds to the analytical approach followed by this thesis. More formally, the overall duration of the relationship between the provider and the customer is defined as T and consists of multiple time segments: $T = \{1, ..., i, ..., n\}$. The subset of time segments in which the customer purchased some products or services from the provider is denoted as U. Using the cardinalities of U and T, the *customer purchase frequency ratio* can be calculated as follows:

$$\text{Customer Purchase Frequency Ratio} = \frac{|U|}{|T|} \qquad (5.37)$$

5.3.3. Proactiveness

In order to assess proactiveness, various empirical measurements have been developed. Wallenburg *et al.* (2010) propose a four-item

scale for measuring proactive improvement behavior of service line employees as perceived by customers in the context of B2B services. This scale helps determining whether provider employees continuously perform process optimization, make suggestions for improvements, adapt the solution to the situation, and take initiatives. Frese *et al.* (1997) associate proactiveness with the degree of personal initiative of provider employees and propose a scale to empirically assess this degree.

Inspired by these approaches, this thesis propose an analytical means to measure the degree of proactiveness by calculating the ratio of proactive improvements to the total number of improvements performed over a specific time period. This information can easily be retrieved from support and service management systems. Support systems indicate the number of actions performed upon problems identified by customers. Service management systems in addition collect the number of change requests performed for each provided solution. More formally, the overall duration of the relationship between the provider and the customer is defined as T and consists of multiple time segments: $T = \{1, ..., i, ..., n\}$. If P_i and R_i represent the number of improvements and changes performed to the solution provided to the customer within the time segment i respectively at the initiative of the provider and of the customer, the *proactiveness ratio* for the time segment i can be calculated as follows:

$$Proactiveness\ Ratio(i) = \frac{P_i}{P_i + R_i} \tag{5.38}$$

5.3.4. Cross-selling

Cross-selling has already been recognized as a key performance indicator in various industries such as in the finance sector (Kamakura *et al.*, 1991), and different approaches have been proposed in order to assess cross-selling achievements from the perspective of the provider. Nash & Sterna-Karwat (1996) propose a methodology to assess cross-selling efficiency based on financial accounts details. Bauer (2004, p.3) elaborates a customer cross-sell index in its KPI profiler

and suggests to calculate this index by "dividing the number of products sold by the number of customers purchasing a product in the last two years." However, these two approaches are not suited in the context of this thesis as they do not allow an individual assessment of the cross-selling performance for each customer. According to Malms & Schmitz (2011, p.258), an effective measure of cross-selling which considers customers on an individual basis has not yet been proposed in past literature: "no prior studies conceptualize or operationalize cross-selling success in a way that accounts for the reliability and validity of the measures." They therefore suggest to measure the effectiveness of the cross-selling activities as the "degree to which the firm exploits customer's full cross-selling potential." However, this assessment is not performed analytically, but empirically using a four item Likert-type scale.

Consequently, inspired by the approach of Malms & Schmitz (2011), this thesis proposes to create a revenue based metric in order to determine the cross-selling performance. Since cross-selling refers to complementing the original offering to the customer with new products and services, this metric is based on the ratio between the revenues generated in a certain time segment by products and services that were already sold to the customer in the past and revenues generated in the same time segment by products and services that the customer purchases for the first time. The overall duration T of the relationship between the provider and the customer is divided in multiple time segments: $T = \{1, ..., i, .., n\}$. P_i represents the set of products and services which are sold to the customer for the first time in the time segment i. Q_i represents the set of products and services which (1) are sold to the customer within i and (2) were sold to the customer before the beginning of i. If $R(P_i)$ and $R(Q_i)$ represent the revenues generated by P_i and Q_i over the time segment i, the *cross-selling revenue ratio* for the time segment i can be calculated as follows:

$$\text{Cross-Selling Revenue Ratio}(i) = \frac{R(P_i)}{R(P_i) + R(Q_i)} \qquad (5.39)$$

Another cross-selling metric based on the cardinalities of P_i and Q_i can also be calculated. This metric, which is defined as *cross-selling diversity ratio* is more qualitative as it only considers the number of products and services contained in P_i and Q_i and it ignores the corresponding revenues:

$$Cross\text{-}Selling\ Diversity\ Ratio(i) = \frac{|P_i|}{|P_i| + |Q_i|} \qquad (5.40)$$

5.3.5. Customer Participation

In order to assess the degree of customer participation, several empirical approaches have been proposed in past literature. Bettencourt (1997) proposes to measure the degree of customer participation with a four item Likert-type scale focusing on the willingness of the customer to share suggestions for improvement and problems. Cermak & File (1994) used a one dimensional construct, asking the actual level of involvement such as invested time and effort to determine customer participation. Inspired by these approaches, this thesis suggests to use the number of proposed improvements by the customer in a given time period in order to assess customer participation. This information can be easily retrieved from support system in which customer issues and requests are stored. More formally, considering the time segment i, and defining as P_i the set of improvements proposed by the customer during i, the metric *customer participation quantity* can be calculated as the cardinality of P_i:

$$Customer\ Participation\ Quantity(i) = |P_i| \qquad (5.41)$$

This metric can be extended with a normalized version called *customer participation ratio* which considers the ratio between the improvements proposed by the customer and the revenues $R(i)$ generated with this customer in the time segment i:

$$Customer\ Participation\ Ratio(i) = \frac{|P_i|}{R(i)} \qquad (5.42)$$

5.3.6. Transaction Costs Reduction

A thorough literature review on existing approaches for measuring transaction costs is proposed by Den Butter (2010, p.15), who argues that "a considerable amount of research must be done" to quantify the transaction costs, because most of the existing research has been theoretical or qualitative. The main approach to assessing transaction costs outlined in Den Butter (2010) is to split the total costs of the provider and the customer in production and transaction costs. In line with this approach, but considering solely the provider's perspective and limiting the scope of the transaction costs analysis to the provider's costs of sales, this thesis proposes to measure the sales related time and resource investments performed by the provider for the customer and to put this value in relationship with the corresponding revenues generated with this customer in the same time period. According to Reichheld & Teal (2001) an organization should invest less time and effort with customers it has established relationships with to generate a certain transaction volume.

The time investments performed by provider employees to identify sales opportunity and transform them into contracts can easily be quantified if the customer-facing activities of the provider employees are tracked. For instance, if the interactions occurring with customer employees such as meetings, phone calls, and letters are stored in the CRM system, the overall customer interaction duration can be measured and corresponds to the previously defined metric volume. This assessment can be extended with the sales related but non-customer facing activities performed by the provider employees such as the preparation of customer meetings, the completion of proofs of concept, and the answering of customer's technical documents. More formally, if I_i represents the total interaction time between sales employees of the provider and customer employees within the time segment i, A_i the total amount of time spent by sales employees of non-customer facing customer-related activity during during the time segment i, and $R(i)$ the revenues generated with the customer during the time segment i, the *transaction effectiveness ratio*

can be calculated as follows:

$$Transaction\ Effectiveness\ Ratio(i) = \frac{I_i + A_i}{R(i)} \qquad (5.43)$$

In summary, chapter 5 elaborated on the *CI Analytics* model and methodology proposed by this thesis to assess customer intimacy. A set of metrics has been conceived in order to assess each of the customer intimacy components proposed in chapter 4 upon available data in the information system of the provider. Eight interaction based metrics have been proposed to assess the acquired knowledge of, and established relationships, with customers at the individual and organizational levels. These metrics are inspired by past literature associating knowledge and relationships to the four interaction characteristics quantity, intensity, regularity, and mode. In addition to these eight metrics, three additional centrality metrics based on the topology of the social network formed by provider and customer employees have been used in order to assess the acquired customer intimacy at the organizational level. Considering the competitive advantages and benefits derived from the customer intimacy strategy, eight metrics based on interaction, activity, and revenue records have been created upon existing literature in order to assess the values of the six leveraged customer intimacy components identified in chapter 4. Furthermore, the *CI Analytics* methodology proposed in section 5.1 allows a calibration of the proposed customer intimacy metrics to the specific interaction patterns of the provider. This methodology is based on the established knowledge discovery in database process (Fayyad *et al.*, 1996a). The next chapter will detail the software implemented in the scope of this thesis to actually calculate these metrics upon real data and to visualize them by means of a graphical user interface.

Part III.
Evaluation

6. *CI Analytics* Software

The software *CI Analytics* has been conceived and implemented in order to validate the *CI Analytics* model and methodology proposed by this thesis in chapter 5. This software enables the calculation of the customer intimacy metrics and makes them available to users by means of a graphical user interface.

This application has been developed in cooperation with the IT software and services provider CAS Software AG, who markets the customer relationship management (CRM) application CAS genesisWorld.[1] More precisely, the software *CI Analytics* accesses the data stored in CAS genesisWorld, such as interaction, project, and revenue data in order to calculate the customer intimacy metrics. Moreover, three students participated in the implementation of the software *CI Analytics* under the supervision of the author of this thesis: Thomas Herzig worked on the elaboration of a first prototype for the calculation of the acquired customer intimacy metrics such as volume, intensity, and frequency of interaction. Johannes Kunze von Bischhoffshausen complemented this prototype with the means to calculate the leveraged customer intimacy metrics. Finally, Hakan Bilgic participated in the analysis of the requirements from the end-

---

[1] Further information on CAS genesisWorld is available at *http://www.cas.de* (accessed on 29.09.2011).

user perspective and in the development of the web-based user interface.

This chapter will elaborate on the technical details of the software *CI Analytics*. Section 6.1 will analyze the business and technical requirements. Section 6.2 will set out the overall architecture of the software *CI Analytics* and illustrate its user interface. Finally, section 6.3 will assess this application with regard to the previously determined requirements and outline the results of a survey on the potential business benefits of this application.

6.1. *CI Analytics* Business Analysis

In order to conceive and implement the software *CI Analytics*, a requirement analysis has been performed. This analysis is developed in section 6.1.1. Subsequently, the relevant business objects for the calculation of the customer intimacy metrics which are contained in the database of the application CAS genesisWorld have been determined and are presented in section 6.1.2.

6.1.1. Requirements Analysis

This section elaborates on the requirements which have been considered for the implementation of the software *CI Analytics*. Following the approach of Sommerville (2007, p.119), this analysis distinguishes functional requirements reflecting the services and behaviors that the system should provide from non-functional requirements which are the constraints on the services provided by the system. In addition, the requirements have been grouped in three distinct domains, each of them covering a specific aspect of the application:

- **Data Source Access**
  This domain covers requirements related to the access to data containing elements of evidence that a certain degree of customer intimacy has been reached between a provider and a customer. According to the *CI Analytics* model proposed in chapter 5 (see figure 5.2), these data sources cover interaction

and activity records, project information as well as financial details on the sales results.

- **Customer Intimacy Calculation**
  This domain covers more specifically the requirements on the calculation of the acquired and leveraged customer intimacy metrics proposed in chapter 5.

- **Customer Intimacy Representation**
  Requirements related to the customer intimacy representation focus on end-users expectations with regard to the graphical user interface (GUI) designed and implemented in the course of this thesis to support the visualization of the customer intimacy metrics.

This requirement analysis has been performed by investigating the behavior of CAS employees with regard to their usage of the application CAS genesisWorld and of their potential usage of the software *CI Analytics*. It considers the end-users perspective as well as the perspectives of the different IT functions that would be responsible for administrating and supporting the software *CI Analytics*. This analysis has led to 15 functional and non-functional requirements which are described in the next parts of this section. Table 6.1 provides a summary of these requirements.

Table 6.1.: Functional and Non-Functional Requirements on *CI Analytics*

No.	Name	Requirement Type	
		Functional	Non-Functional
Data Source Access			
1	Access and process data stored in the application CAS genesisWorld	✓	
2	Support the access to additional data sources	✓	

Functional and Non-Functional Requirements on *CI Analytics* (continued)

No.	Name	Requirement Type	
		Functional	Non-Functional
3	Minimize performance impact on CAS genesisWorld		✓
4	Provide scalable algorithm to access the data		✓
5	Ensure that sensitive data is securely handled		✓
Customer Intimacy Calculation			
6	Consider calibration parameters to perform the metrics calculation	✓	
7	Calculate the acquired customer intimacy metrics at the individual level	✓	
8	Calculate the acquired customer intimacy at the organizational level, including the network based centrality metrics	✓	
9	Calculate the leveraged customer intimacy metrics	✓	
10	Use efficient algorithms and scalable architecture to calculate the customer intimacy metrics		✓
11	Incrementally update the customer intimacy metrics values		✓
Customer Intimacy Representation			
12	Visualize the acquired customer intimacy metrics at the individual level by means of a graph representation	✓	
13	Visualize the acquired customer intimacy metrics at the organizational level	✓	
14	Visualize the leveraged customer intimacy metrics by means of charts	✓	

Functional and Non-Functional Requirements on *CI Analytics* (continued)

No.	Name	Requirement Type	
		Functional	Non-Functional
15	Provide data visualization by means of a web-based interface	✓	

6.1.1.1. Data Source Access

The following functional requirements related to the access to data sources have been determined:

1. **Access and process data stored in the application CAS genesisWorld**
   In order to validate the proposed approach of this thesis, the customer intimacy metrics are calculated upon the data stored in the application CAS genesisWorld. Thus, the software *CI Analytics* should be able to access the underlying database of CAS genesisWorld and process its data in a way that enables the calculation of the customer intimacy metrics.

2. **Support the access to additional data sources**
   Even though the software *CI Analytics* primarily focuses on data contained in CAS genesisWorld, its architecture should provide the ability to easily incorporate other sources of data such as other CRM systems, project databases, groupware, or social platforms. Thus, the architecture of the software *CI Analytics* should be structured in a way that the data integration is separated from the actual calculation of the customer intimacy metrics.

In addition, the following non-functional requirements have been determined:

3. **Minimize performance impact on CAS genesisWorld**
   CAS genesisWorld is a business critical application used by all customer-facing employees such as sales representatives, services employees, and project managers. It is therefore mandatory that the software *CI Analytics* does not significantly impact the performance of CAS genesisWorld: the access to the data and the metrics calculation should be transparent to the end-users of CAS genesisWorld. Thus, the connections between *CI Analytics* and CAS genesisWorld should be minimized and efficiently performed. It should be possible to complete the resource-intensive customer intimacy calculation on a separate computer, and the resulting customer intimacy metrics should be stored in their own database, externally to the CAS genesisWorld database.

4. **Provide scalable algorithm to access the data**
   In order to calculate the customer intimacy metrics for a specific customer, the software *CI Analytics* must retrieve all interactions such as emails, meetings, phone calls, projects activities, and sales results related to this customer and stored in CAS genesisWorld. Considering the potentially high number of employees involved in the relationship between the provider and customer, and the duration of this relationship which may span over several years, these records can lead to multiple gigabytes of data. The algorithms for retrieving the data stored in CAS genesisWorld must, therefore, be efficient and scalable in order to handle large amounts of data.

5. **Ensure that sensitive data is securely handled**
   CAS genesisWorld contains sensitive customer related information such as project information, sales results, and specific interaction records between provider and customer employees. This information must be carefully managed to ensure that the data is solely used for the purpose of calculating the customer intimacy metrics. In addition the data access rights specified in CAS genesisWorld should be propagated to the software *CI Analytics* to make sure that the data is only accessed with the appropriate credentials.

6.1.1.2. Customer Intimacy Calculation

The following functional requirements related to the calculation of the customer intimacy metrics have been identified:

6. **Consider calibration parameters to perform the metrics calculation**

   Multiple parameters have been determined in chapter 5 in order to enable a calibration of the *CI Analytics* model to the specific patterns of each provider. These parameters are the time period, the segment size, the interaction duration threshold, the interaction quantity threshold, and the weighted interaction quantity threshold. The software *CI Analytics* should provide the ability to specify the values of these parameters as well as to consider them in the calculation of the customer intimacy metrics.

7. **Calculate the acquired customer intimacy metrics at the individual level**

   In chapter 5, eight metrics have been conceived upon the concepts of customer interaction time and weighted customer interaction time in order to determine the acquired customer intimacy at the individual level. These metrics are for instance volume, intensity, and frequency of interaction. Thus, the software *CI Analytics* should be able to (i) retrieve all provider employees involved in the relationship with a specific customer, (ii) retrieve all customer employees involved in this relationship, (iii) calculate the customer interaction time and weighted customer interaction time for each provider-customer employee combination, and (iv) calculate the eight corresponding customer intimacy metrics for each of these combinations.

8. **Calculate the acquired customer intimacy metrics at the organizational level, including the network-based centrality metrics**

   The software *CI Analytics* should be able to aggregate the customer interaction time and weighted customer interaction time calculated at the individual level in order to determine the values of the customer intimacy metrics at the organizational

level. In addition, the software *CI Analytics* should include a graph technology allowing the calculation of the centrality metrics degree centrality, normalized degree centrality, and normalized closeness centrality.

9. **Calculate the leveraged customer intimacy metrics**
   Eight metrics have been proposed in chapter 5 to assess the leveraged customer intimacy components such as the customization revenue ratio, the customer purchase frequency ratio, and the cross-selling revenue ratio. The software *CI Analytics* should be able to determine the values of these customer intimacy metrics for all customers of the provider for different time frames such as the past quarter or the past year.

The non-functional requirements related to the calculation of the customer intimacy metrics are the following:

10. **Use efficient algorithms and scalable architecture to calculate the customer intimacy metrics**
    A high quantity of data, up to multiple thousands of interaction, project, and sales records has to be processed and evaluated in order to calculate the customer intimacy metrics at the individual and organizational levels. Therefore, *CI Analytics* must use efficient algorithms and rely on a scalable architecture in order to process this data and calculate the customer intimacy metrics.

11. **Incrementally update the customer intimacy metrics values**
    In order to take advantage of the proposed customer intimacy metrics, this information must be precise and up-to-date. Therefore, the metrics should be automatically recalculated on a periodical basis, taking into account the most recent data such as the last emails or the sales achievements stored in CAS genesisWorld. The calculation frequency impacts the number of access to the data sources and, therefore, should be configurable. For instance, the calculation of the customer intimacy metrics could occur on a daily, weekly, or monthly basis.

6.1.1.3. Customer Intimacy Representation

The following functional requirements have been established with regard to the representation and visualization of the customer intimacy information:

12. **Visualize the acquired customer intimacy metrics at the individual level by means of a graph representation**
    In order to graphically depict the values of the acquired customer intimacy metrics at the individual level, the software *CI Analytics* should provide a graph based representation of all relationships established by provider employees with customer employees, as illustrated in figure 1.1. In this graph, the nodes should represent the provider and customer employees, and the edges should reflect the customer intimacy established between the corresponding employees. In addition, *CI Analytics* should provide the ability to specify the customer intimacy metric used to determine the weights of the edges on the graph as well as the considered time period for the calculation of the metrics.

13. **Visualize the acquired customer intimacy metrics at the organizational level**
    The software *CI Analytics* should provide the ability to visualize the eight acquired customer intimacy metrics at the organizational level.

14. **Visualize the leveraged customer intimacy metrics by means of charts**
    The software *CI Analytics* should provide the ability to visualize the eight leveraged customer intimacy metrics in the form of column or line charts, thereby representing the evolution of the metrics over time.

15. **Provide data visualization by means of a web-based interface**
    In order to access the calculated customer intimacy metrics, a web-based interface should be provided which allows a remote access via Internet to the information. This interface should include all information related to the acquired and leveraged

customer intimacy and its development should adhere to technology standards. It should also provide the ability to select a customer in a list and to represent the customer intimacy information for this specific customer only. It should in addition provide the ability to filter and sort the displayed information.

Since the software *CI Analytics* is in its current state a prototypical implementation serving research purposes, the following non-functional requirements are out of the scope of this thesis: allowing a customization of the interface, implementing an authentication mechanism to access the application, and guarantying specific service quality levels such as response time and availability.

6.1.2. Business Objects Analysis

Even though future versions of the software *CI Analytics* should provide the ability to retrieve data from various sources containing relevant information for the calculation of the customer intimacy metrics, the current version of *CI Analytics* focuses on the data contained in the application CAS genesisWorld. This section outlines the data retrieved by the software *CI Analytics* from the application CAS genesisWorld in order to calculate the customer intimacy metrics.

In CAS genesisWorld, the business objects of type *Address*, represent either a customer organization, a customer employee, or a provider employee. This is a central item in the architecture of CAS genesisWorld as it contains all customer related details such as names and addresses. In addition to the business objects of type *Address*, the analysis of the CAS genesisWorld database has led to the identification of nine relevant business objects in the context of this thesis. These nine business objects are presented in table 6.2 and can be categorized in three main categories:

- **Interaction Business Objects** record the interactions that occurred between the provider and customer employees. They are required to calculate the customer interaction time and weighted customer interaction time which in turn are used to calculate the acquired customer intimacy metrics such as volume, intensity, and frequency of interaction.

- **Activity Business Objects** record the activities performed by the provider employees. They are required as input to the leveraged customer intimacy metrics to assess the time spent by provider employees on customer projects and on the resolution of customer problems.

- **Revenue Business Objects** track the details on the monetary and non-monetary revenue generated with customers. The monetary revenue reflects the sales transaction achievements and the non-monetary revenue concerns other form of value provided by the customer such as the customer's suggestions for improvements.

Table 6.2.: *CI Analytics* Business Objects

No.	Type	Description
Address Business Objects		
1	Address	Addresses represent customer organizations, customer employees, and provider employees in the application CAS genesisWorld. Each address record contains the required contact information such as name, addresses, phone numbers, as well as his preferences, his preferred contact, and links to past activities.
Interaction Business Objects		
2	Email	Exchanged emails with the customer can be automatically stored in CAS genesisWorld and, thus, are available for the calculation of customer intimacy metrics. Details such as the sender and receivers of the emails, time stamps, and content can be retrieved.
3	Appointment	Appointments consist of the meetings organized with the customer. They can be entered directly in CAS genesisWorld or retrieved from calendar applications. Details such as the list of participants to meetings as well as the date and duration of meetings can be retrieved.

CI Analytics Business Objects (continued)

No.	Type	Description
4	Phone Call	CAS genesisWorld can integrate a business phone system, thereby allowing to store details on the phone calls happening between provider and customer employees, such as the provider and customer employees names (through a mapping based on their phone numbers), the phone calls dates and durations.
5	Document	Documents refer to the letters received from, and sent to, customer employees. Sender and receivers names as well as the document date are available in each document record.
Activity Business Objects		
6	Project Activity	Customer projects are decomposed in multiple project activities which can be accessed in order to assess the tasks performed and time spent by provider employee on customer projects. The activity type, list of involved provider employees, date, and duration are available in each project activity record.
7	Service Ticket	When a customer has a specific request or if he experiences an issue with the provided solution, a service ticket is created and managed by the support team until its closure. Service tickets are, thus, indicators of the efforts performed by the provider to support the customer. The names of the customer and involved customer and provider employees as well as the date, the time spent, and the actions undertaken to solve the problem are available in each service ticket record.
Revenue Business Objects		
8	Invoice Line Item	Invoice Line Items provide details on the products and services purchased by the customer, such as the product and service references, the price paid, and the date of purchase.

	CI Analytics Business Objects (continued)	
No.	Type	Description
9	Suggestion	Suggestions are special types of service tickets in which customers report suggestions for improvement. They are thus, to some extent, indicators of the degree of participation of the customer to the development and improvement of the solution.

6.2. *CI Analytics* Architecture

6.2.1. Architecture Overview

This section presents an overview of the architecture of the software *CI Analytics*. This architecture relies on a data warehouse and on an *extract, transform, load (ETL)* component for extracting and processing the data stored in operational databases. It adheres, thus, to the architectural standards for decision support applications proposed by Turban *et al.* (2011). Figure 6.1 depicts the main components of this architecture as well as the interface of the software *CI Analytics* with an external customer data source such as CAS genesisWorld. This architecture is structured along the three following layers:

- **Data Layer (*CI Data Warehouse*)**
  This layer contains the *CI Data Warehouse*, which is the underlying database of the software *CI Analytics*, as well as the operational database of the considered source of customer intimacy to calculate the customer intimacy metrics, in particular, the database of the application CAS genesisWorld. *CI Data Warehouse*, as suggested by its name, is an implementation of a data warehouse. Turban *et al.* (2011, p.328) define a data warehouse as a "repository of current and historical data of potential interest to managers throughout the organization" and precise that its data is structured in a way that supports analytical processing activities. *CI Data Warehouse* is not used for operational

purposes, but solely stores customer related data which is relevant for the calculation of the customer intimacy metrics. It is also optimized for efficiently reading and processing large amounts of data. Further details on the *CI Data Warehouse* are provided in section 6.2.2.

- **Application Layer (*CI ETL* and *CI Services*)**
  This layer reflects the business logic components of the software *CI Analytics*, namely the *CI ETL* and the *CI Services*:

  – The component *CI ETL* populates the database *CI Data Warehouse* based on the data available in operational databases such as the CAS genesisWorld database. *CI ETL* implements an *extract, transform, and load* process (Turban *et al.*, 2011, p.344). It reads the data from the operational database, filters the relevant records, transforms them to prepare the calculation of the customer intimacy metrics, and loads the transformed data in the *CI Data Warehouse* database. This component is further developed in section 6.2.3

  – The *CI Services* are the clients of the database *CI Data Warehouse*. They expose the functionality of calculating the different customer intimacy metrics upon the data available in *CI Data Warehouse* by means of RESTful web services. RESTful web services use the standard http protocol and adhere to the REST – Representational State Transfer – architecture which simplifies components interoperability, increases scalability, and provides an easy access to the customer intimacy metrics (Richardson & Ruby, 2007). Section 6.2.4 elaborates on the *CI Services*.

- **Presentation Layer (*CI Dashboard*)**
  This layer focuses on the presentation of the customer intimacy information to the users and consists of the component *CI Dashboard*. The *CI Dashboard* is the graphical user interface (GUI) implemented in the software *CI Analytics*. It provides a graph-based representation of the acquired customer intimacy and a chart-based representation of the leveraged customer intimacy.

Users can input the necessary parameters such as customer name, time frame and requested metrics in order to configure the *CI Services* calls. The *CI Dashboard* is a web-based graphical user interface, letting users access the information with their web browser from their organization's intranet or from Internet. Further information on this component is proposed in section 6.2.5.
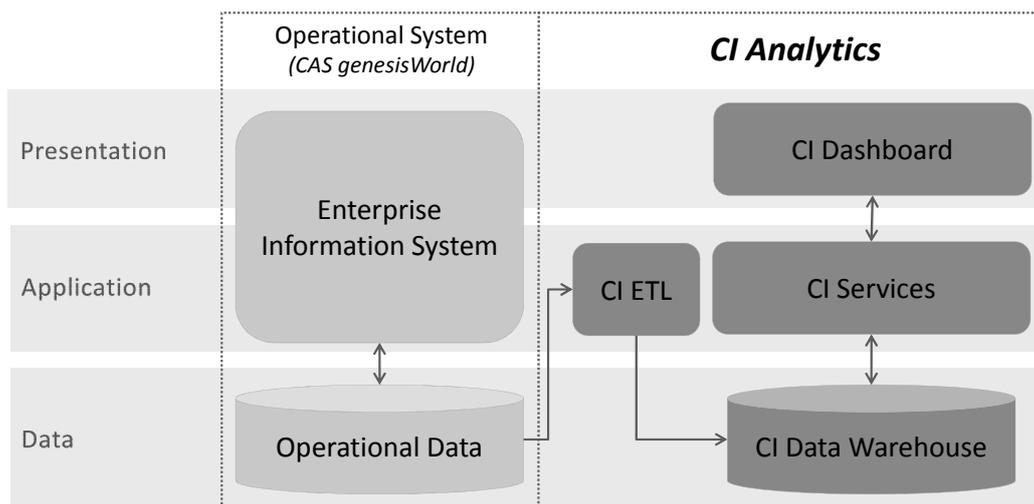
Figure 6.1.: *CI Analytics* Architecture

The next sections detail each of the previously mentioned components of the software *CI Analytics*.

6.2.2. CI Data Warehouse

The *CI Data Warehouse* is the underlying database of the software *CI Analytics*. It stores data in a form allowing the calculation of the customer intimacy metrics after it has been extracted from the different operational sources of customer intimacy data, such as the operational database of the application CAS genesisWorld. A key aspect of data warehouses is that they are subject-oriented and multidimensional (Turban *et al.*, 2011, p.332). This means that the data is organized along specific subjects and it is structured in a way that supports its analysis along multiple dimensions. In order to fulfill these characteristics, the design of data warehouse tables follows the

star schema which consists of a central subject-focused fact table surrounded by multiple dimension tables (Turban *et al.*, 2011, p.351).

Following this approach, the fact tables in the *CI Data Warehouse* focus on the key elements required to perform the calculation of the customer intimacy metrics. Three fact tables have been conceived:

- **Customer interaction time fact table**
  The customer interaction time fact table stores details on the interaction time spent with each customer employee. Each interaction business object stored in CAS genesisWorld such as emails, meetings, phone calls, and documents are transformed by the *CI ETL* process into a record in the customer interaction time fact table. This record contains the duration of the interaction act to allow a calculation of customer interaction time and weighted customer interaction time as well as additional dimensional information. In order to provide multiple dimensional analyses of the customer interaction time, the following dimension tables have been conceived:

  - *CustomerCompany*, to focus on a specific customer organization,

  - *CustomerEmployee*, to focus on a specific customer employee,

  - *ProviderEmployee*, to focus on a specific provider employee,

  - *Date*, to focus on a specific time frame,

  - *Channel*, to focus on a specific interaction channel,

  - *Project*, to focus on a specific project.

  Figure 6.2 illustrates the customer interaction time fact table surrounded by its six dimension tables. Using these seven tables it is possible to combine multiple pieces of dimensional information to calculate the customer interaction time, weighted customer interaction time, and subsequently the acquired customer intimacy metrics for very specific criteria. For instance, it is possible to calculate the acquired customer intimacy metrics for the provider employee p with the customer employee

c from the customer organization *C* within a specific year. This analysis could be further detailed by specifying a channel of interaction or a project reference.
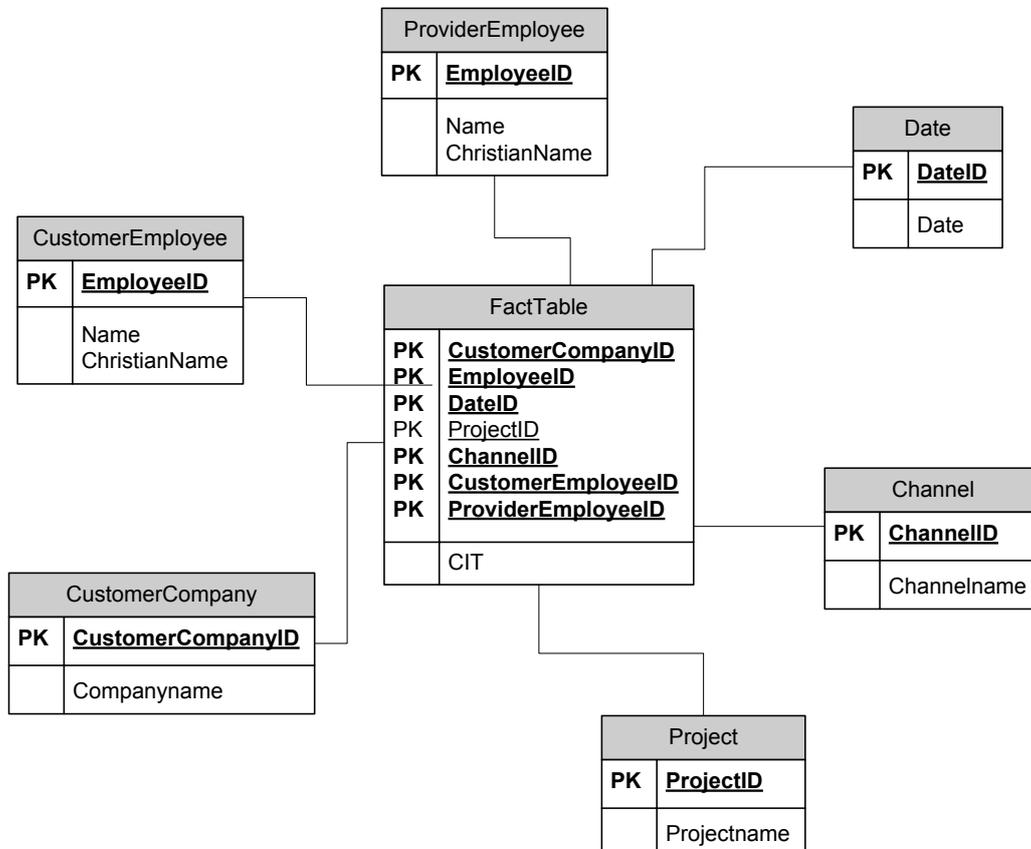
Figure 6.2.: Customer Interaction Time Star Schema

- **Customer activity time fact table**
  The customer activity time fact table focuses on the duration of the activities performed by provider employees for customers. Similarly to the interaction business objects, the business objects of type project activity or service ticket are transformed by the *CI ETL* process into records in the customer activity time fact table. Records in this table contain the activity duration as well as dimensional values corresponding to the six dimensional tables surrounding the customer activity time fact table. Five of these dimension tables are the same as those of the customer interaction time fact table, namely: *CustomerCompany*,

CustomerEmployee, *ProviderEmployee*, *Date*, and *Project*. However, the dimension table *ActivityType* replaces the dimension table *Channel*. The dimension *ActivityType* allows to evaluate whether the activity is value-adding like a consulting task, or non-value-adding such as an administrative task.

- **Customer value return fact table**
  Records in the customer value return fact table contain details on the monetary and non-monetary revenues generated with the different customers of the provider. Business objects of type invoice line item represent the monetary revenues generated with customers such as the achieved sales transactions. They are, thus, converted into records containing their monetary value in the customer value return fact table. Business objects of type suggestion represent a special form of non-monetary customer value return as the information provided in the suggestion can be used by the provider to improve its value proposition, thereby enabling him to achieve a new competitive advantage. Therefore, the business objects of type suggestion are also transformed into records in the customer value return fact table. In the current version of the software *CI Analytics*, the business objects of type suggestion are converted into records having a constant monetary value. Its architecture, however, would support a monetary quantification of the customer suggestions which could be elaborated in future research. The dimensional information provided in the records of the customer value return fact table allows to distinguish monetary revenues derived from invoice line items from the non-monetary revenues which are derived from suggestions. It also allows the calculation of the leveraged customer intimacy metrics for specific customers, time periods, or projects. Four dimension tables therefore surround the customer value return fact table:

  - *CustomerCompany*, to focus on the revenues generated with a specific customer organization,

  - *Project*, to focus on the revenues derived from a specific project,

- *Date*, to estimate the revenues in a specific time frame,

- *ValueSource* to determine whether the record refers to monetary or non-monetary revenue.

In order to implement the *CI Data Warehouse*, the application Microsoft SQL Server 2008 R2 has been used.[2] This application was chosen because it provides the required tools to realize a data warehouse upon standard database management functions, thereby decreasing the complexity of the overall software architecture, and because this is the default database of the application CAS genesisWorld.

6.2.3. CI ETL

The component *CI ETL* implements the extract, transform, and load process of the software *CI Analytics*. It is responsible for populating the database *CI Data Warehouse*. In the extraction phase, data which is relevant for the calculation of the customer intimacy metrics is read out of the operational databases, such as the database of CAS genesisWorld. During the transformation phase, this data is filtered and converted into the format of the *CI Data Warehouse* in order to be entered in one of the fact tables or one of the dimension tables. Finally, during the loading phase, the transformed data is actually stored in the fact and dimension tables of the *CI Data Warehouse*. The process of the *CI ETL* component consists of eight subprocesses which are depicted in figure 6.3:

1. *ETL CustomerCompany Data* extracts customer organizations details stored in the business objects of type address such as the names and reference numbers of the companies. It subsequently transforms and loads them in the dimension table *CompanyName*.

2. *ETL CustomerEmployee Data* extracts data related to customer employees which is stored in the business objects of type address. Then, it transforms and loads it in the dimension table *CustomerEmployee*.

---

[2] Further details are available at
http://www.microsoft.com/sqlserver/en/us/default.aspx (accessed on 29.09.2011).

3. *ETL ProviderEmployee Data* extracts details on provider employees, which are also stored in the form of business objects of type address. These objects are subsequently transformed and loaded in the dimension table *ProviderEmployee*.

4. *ETL Project Data* extracts details on customer projects which are stored in the business objects of type project activity. It then transforms and loads them in the dimension table *Project*.

5. *ETL Activity Data* extracts details on the activity durations from the business objects of type project activity or service ticket. This information is then used to populate the customer activity time fact table .

6. *ETL Revenue Data* extracts financial information out of the business objects of type invoice line item. Subsequently, it transforms this data and loads it as facts into the customer value return fact table.

7. *ETL Customer Participation Data* filters the business objects of type service request which are specifically referring to customer suggestions, then transforms and loads them as facts in the customer value return fact table. In the current version, as explained in section 6.2.2, these specific facts all have the same monetary value and can be differentiated from the monetary customer revenues derived from business objects of type invoice line items through the *ValueSource* dimension.

8. *ETL Interaction Data* extracts the interaction duration as well as the required dimensional information from the interaction business objects of type email, appointment, phone call, or document. Then, it transforms this data and loads it into the customer interaction time fact table.

The implementation of the component *CI ETL* has been performed with the Microsoft SQL Server Integration Services since both the *CI Data Warehouse* and the database of CAS genesisWorld are realized with Microsoft SQL Server 2008 R2.[3]

---

[3]   Further details are available at
       http://www.microsoft.com/sqlserver/en/us/solutions-technologies/business-

CAS genesisWorld Business Objects
- Address
- Project Activity
- Service Ticket / Suggestion
- Invoice Line Item
- Email
- Appointment
- Phone Call
- Document

CI Data Warehouse

Dimension Tables
- CustomerCompany
- CustomerEmployee
- ProviderEmployee
- Project

Fact Tables
- Customer Activity Time
- Customer Revenue Value
- Customer Interaction Time

1.	ETL Customer Company Data	5.	ETL Activity Data
2.	ETL Customer Employee Data	6.	ETL Revenue Data
3.	ETL Provider Employee Data	7.	ETL Participation Data
4.	ETL Project Data	8.	ETL Interaction Data

Figure 6.3.: Overview of the CI ETL Process

6.2.4. CI Services

The *CI Services* provide client applications such as the *CI Dashboard* with an access to the data stored in the *CI Data Warehouse* and expose the functionality of calculating the customer intimacy metrics by means of standardized RESTful web services. They thereby enable the calculation of the acquired and leveraged customer intimacy metrics proposed in chapter 5.

Following the RESTful web services approach, client applications invoke the *CI Services* with a *GET* request containing the required input parameters such as the name of the metric and the considered time frame to calculate the customer intimacy metrics. The *CI Services*

_____

intelligence/integration-services.aspx (accessed on 29.09.2011).

subsequently convert these requests into a set of SQL queries, perform these queries on the data contained in the *CI Data Warehouse*, aggregate the results, perform the calculation of the customer intimacy metrics, and finally return the results to the client application in an XML format. Technical details such as the input and output parameters of the *CI Services* are presented in appendix E.

Each *CI Service* in the software *CI Analytics* is designed to calculate one of the acquired and leveraged customer intimacy metrics proposed in chapter 5. Thus, the following *CI Services* have been conceived and implemented:

- *CI Services* **for the Acquired Customer Intimacy Metrics**
  In section 5.2, this thesis establishes eight metrics to assess the acquired customer intimacy upon the concept of customer interaction time and weighted customer interaction time, namely volume, weighted volume, intensity, weighted intensity, frequency, duration, number of episodes, and mode of interaction. Moreover, two levels of analysis have been proposed:

  - The *individual* level of analysis allows an assessment of the degree of customer intimacy established between provider and customer employees. Consequently, eight services have been conceived to assess the acquired customer intimacy metrics at the individual level. These services take as input the reference to a customer organization, the beginning and end dates of the chosen calculation time period, and the calibration parameter values specified in appendix E.1. They return a graph in the DyNetML format proposed by Tsvetovat *et al.* (2004). This graph represents the social network formed by the provider and customer employees: its nodes symbolize the employees, and the weights on the edges of the graph are the actual customer intimacy metrics values.

  - The *organizational* level of analysis considers the degree of customer intimacy established between a provider employee and a customer organization. Consequently, eight

services have been realized in order to implement the calculation of the eight acquired customer intimacy metrics at the organizational level. These services take as inputs a reference to a customer organization, the beginning and end dates of the chosen calculation time period, different calibration parameters which are specified in appendix E.1 as well as the reference to the provider employee for which the metric is calculated. These *CI Services* return the values of the considered customer intimacy metrics. These eight services are presented in table 6.3 and further detailed in appendix E.1. At the organizational level of analysis, in addition to the eight customer interaction time based metrics, three network centrality metrics have been conceived, namely the number of contacts (degree centrality), the normalized degree centrality, and the normalized closeness centrality. These services have not yet been implemented in the software *CI Analytics*. However, they have been implemented in the first *CI Analytics* prototype called *CI Graph*. Appendix E.3 provides additional details on *CI Graph*.

- *CI Services* **for the Leveraged Customer Intimacy Metrics**
  Eight customer intimacy metrics have been proposed in section 5.3 in order to assess the leveraged customer intimacy components. These metrics are: customization revenue ratio, customer purchase frequency ratio, proactiveness ratio, cross-selling revenue ratio, cross-selling diversity ratio, customer participation quantity, customer participation ratio, and transaction effectiveness ratio. With the exception of the metric proactiveness ratio for which no data is available in the application CAS genesisWorld, all leveraged customer intimacy metrics have been implemented in a *CI Service*. Thus, seven services have been realized in the software *CI Analytics*. Table 6.3 presents these seven services and the corresponding customer intimacy metrics. Further details such as the inputs and outputs of the services are provided in appendix E.2.

The *Windows Communication Foundation*[4] which is part of the Microsoft .NET framework has been used in order to implement the *CI Services* as resource-oriented REST services (Chappell, 2010). This technology has been chosen because it provides an easy integration with the underlying database of the *CI Date Warehouse* Microsoft SQL Server 2008 R2. Since the created services are available via the standard http protocol, these services are not constrained into the .NET environment but can be accessed by any application supporting the http protocol. The actual development has been performed with the software Microsoft Visual Studio in the programming language C#.[5]

Table 6.3.: *CI Services* Overview

Customer Intimacy Metric	CI Service	
	Individual Level	Organizational Level
Acquired Customer Intimacy		
Volume	Volume Service	Org Volume Service
Weighted Volume	WVolume Service	Org WVolume Service
Intensity	Intensity Service	Org Intensity Service
Weighted Intensity	WIntensity Service	Org WIntensity Service
Frequency	Frequency Service	Org Frequency Service
Duration	Duration Service	Org Duration Service
Number of Episodes	NumberEpisodes Service	Org NumberEpisodes Service
Mode	Mode Service	Org Mode Service
Number of Contacts (Degree Centrality)		Available in first prototype only
Normalized Degree Centrality		Available in first prototype only

---

[4]   Further details are available at
       http://msdn.microsoft.com/en-us/netframework/aa663324 (accessed on 29.09.2011).
[5]   Further details are available at
       http://msdn.microsoft.com/en-us/vcsharp/aa336809 (accessed on 29.09.2011).

Table 6.3.: *CI Services* Overview (Continued)

Customer Intimacy Metric	CI Service	
	Individual Level	Organizational Level
Normalized Closeness Centrality		Available in first prototype only
Leveraged Customer Intimacy		
Customization Revenue Ratio		Customization Revenue Ratio Service
Customer Purchase Frequency Ratio		Customer Purchase Frequency Ratio Service
Proactiveness Ratio		
Cross Selling Revenue Ratio		CrossSelling Revenue Ratio Service
Cross-Selling Diversity Ratio		CrossSelling Diversity Ratio Service
Customer Participation Quantity		Customer Participation Quantity Service
Customer Participation Ratio		Customization Participation Ratio Service
Transaction Effectiveness Ratio		Transaction Effectiveness Ratio Service

6.2.5. CI Dashboard

The *CI Dashboard* provides the means to graphically visualize the acquired and leveraged customer intimacy metrics. In order to facilitate its adoption, the *CI Dashboard* has been implemented as a web-application which is accessible with an Internet browser. Figure 6.4 illustrates the main interface of the *CI Dashboard* with fictive data.
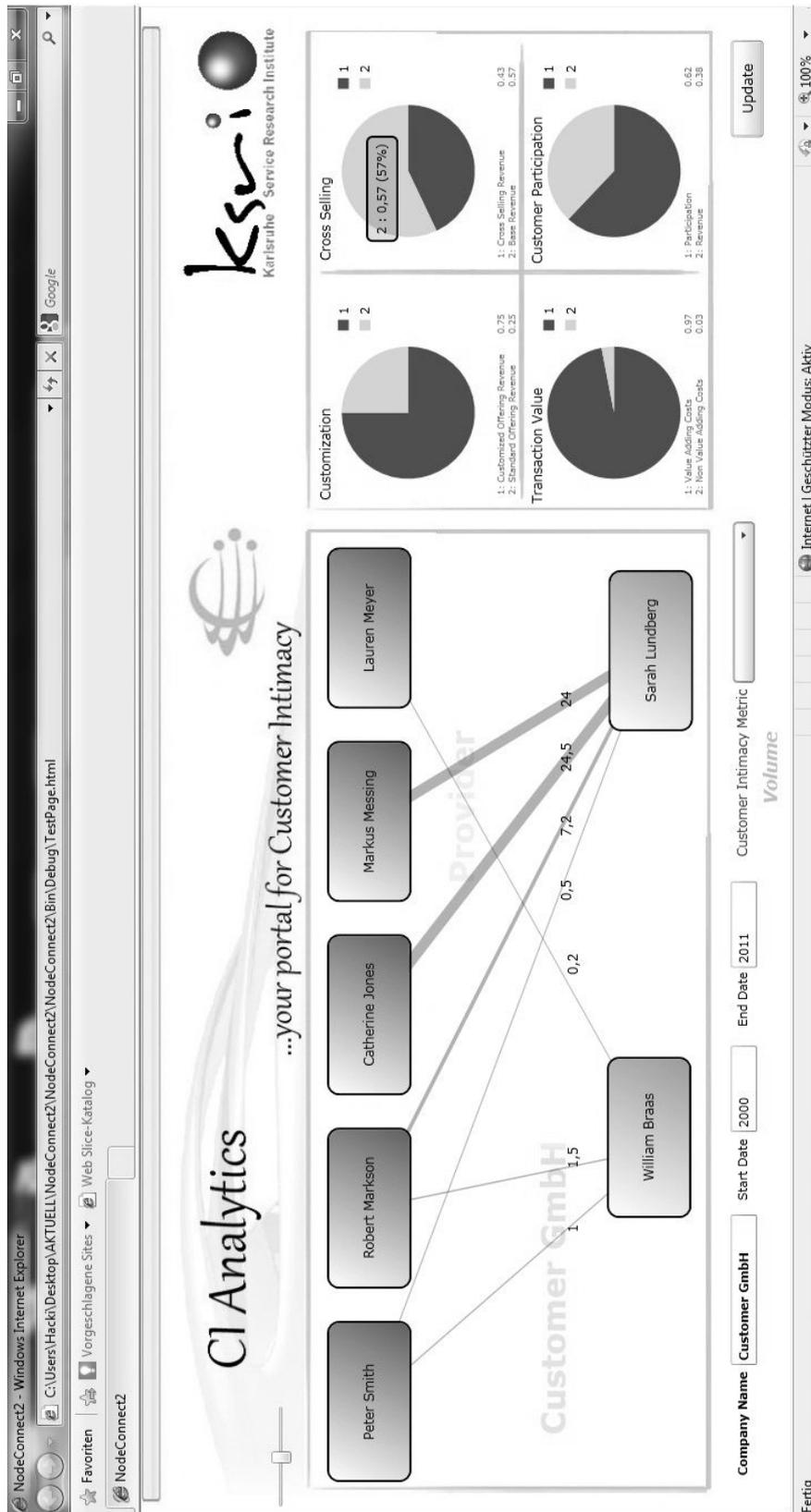
Figure 6.4.: Main Interface of the CI Dashboard

The bottom part of the interface allows users to specify the name of a customer, the first and last years of the considered time period, and the metric to be displayed on the edges of the social network representation (volume in figure 6.4). After clicking on the *Update* button located in the bottom right corner, web services are called in order to render the two parts of the *CI Dashboard* reflecting the assessment of the acquired and leveraged customer intimacy:

- **Acquired Customer Intimacy Visualization**
  The diagram on the left-hand part of the *CI Dashboard* provides a representation of the social network formed by the provider and customer employees. The rectangles symbolize the employees and are aligned in two rows: the rectangles in the top row represent the provider employee and those in the bottom row represent the customer employees. The edges on the diagram connect the provider employees to the customer employees. An edge between the provider employee p and the customer employee c indicates that the chosen acquired customer intimacy metric calculated for the couple of employees $\{p, c\}$ on the specified time frame has a value greater than 0. The weights of the edges which are displayed on the diagram indicate the actual values of the selected customer intimacy metric. For instance, the metric volume has a value of 24.5 for the couple of employees {"Catherine Jones" ; "Sarah Lundberg"}. This means that during the specified time period, the provider employee "Catherine Jones" interacted for a duration of 24.5 hours with the customer employee "Sarah Lundberg".

  The *CI Dashboard* provides the functionality to zoom into the diagram by selecting an employee and using the slider on the top left corner. Moreover when an employee is selected, a new rectangle is displayed, as illustrated in figure 6.5(a), in which detailed information on the provider employee can be displayed such as his role and organization units. In the future, the *CI Dashboard* should provide the ability to display the acquired customer intimacy metrics at the organizational level in this rectangle. Figure 6.5(b) illustrates the capability of the software

CI Analytics and especially of the *CI Dashboard* to handle large amount of data, thereby demonstrating the scalability of the implemented algorithms. This diagram is based upon real data which has been anonymized. This picture also emphasizes the needs for additional filtering capabilities which should be provided in a future version.

- **Leveraged Customer Intimacy Visualization**
  The right-hand side of the *CI Dashboard* illustrated in figure 6.4 provides the functionality to visualize the leveraged customer intimacy metrics by means of chart diagrams. The current version of *CI Dashboard* allows a representation of three out of the eight proposed customer intimacy metrics, namely customization revenue ratio, cross-selling revenue ratio, and customer participation ratio. The *CI Dashboard* will be extended in the next release to include the remaining leveraged customer intimacy metrics such as the customer purchase frequency ratio or the transaction effectiveness ratio.

  While the default representation of the customer intimacy metrics uses pie charts, column charts can also be displayed, as illustrated in figure 6.6 for the metric cross-selling revenue ratio. Using this representation, it is possible to analyze the evolution of customer intimacy metric value over time.

The *CI Dashboard* has been implemented with the technology Silverlight.[6] Silverlight is an application framework enabling the creation and delivery of rich internet applications which can be installed as a plug-in in the Internet Browser. This technology has been chosen because it provides a good integration with the *Windows Communication Foundation* used to implement the *CI Services* and it is a technology used in the latest release of CAS genesisWorld.

In order to realize the graph representation supporting the visualization of the acquired customer intimacy metrics at the individual level, the technology NodeConnect developed by Hodnick (2009) has

---

[6]   Further details are available at *http://www.microsoft.com/silverlight* (accessed on 24.09.2011).
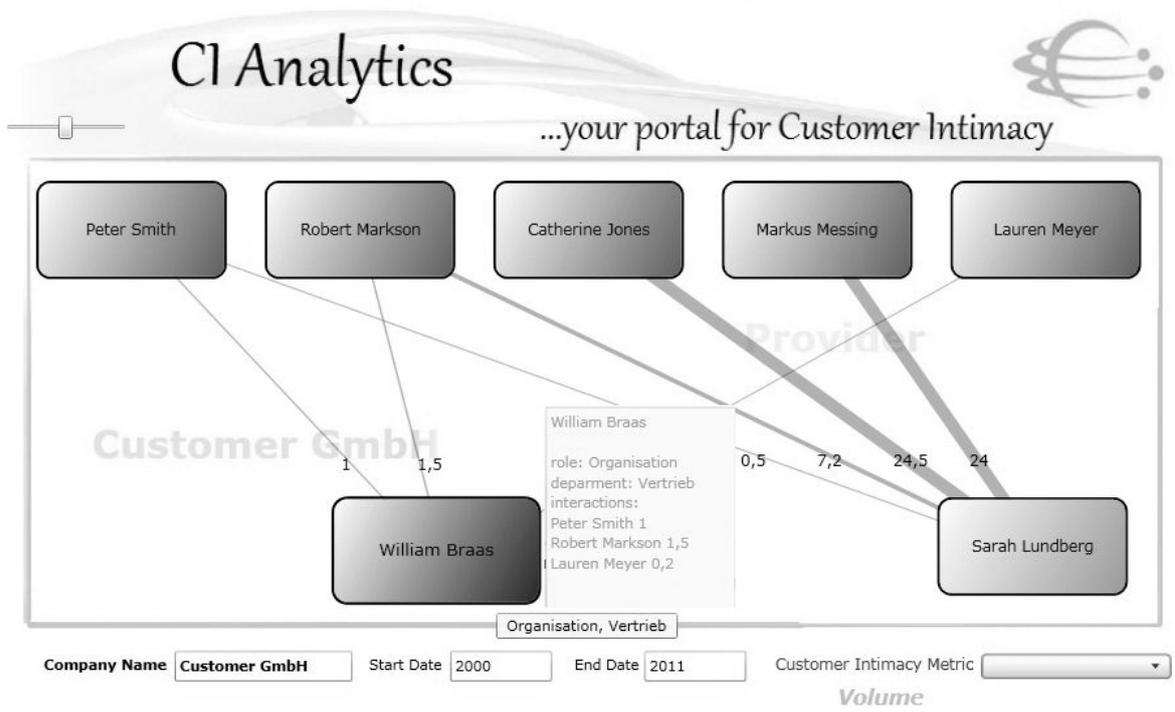
(a) Detailed Information
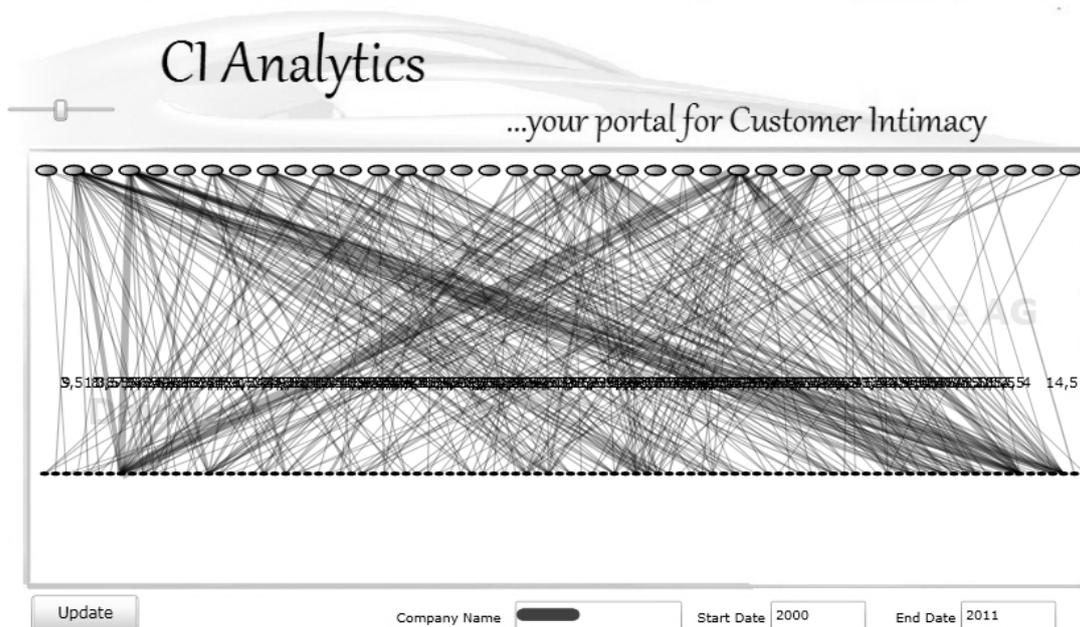
(b) Large Social Network

Figure 6.5.: CI Dashboard: Acquired Customer Intimacy
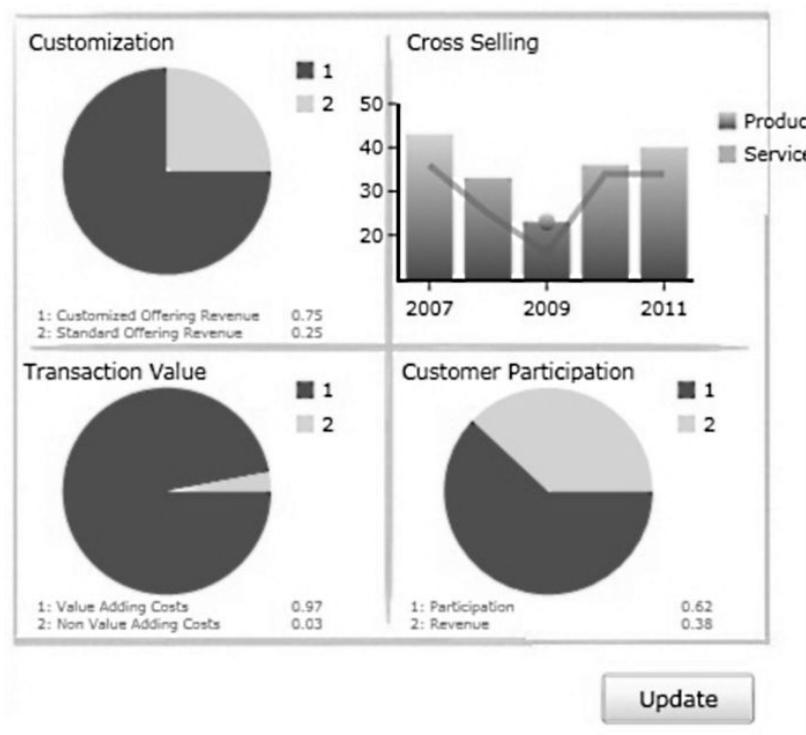
Figure 6.6.: CI Dashboard: Leveraged Customer Intimacy

been chosen as it consists of a simple and free of charge library allowing an easy customization of the graph representation. In order to realize the chart-based representation of the leveraged customer intimacy metrics, the Quickchart library developed by amCharts[7] has been chosen as this technology allows to fulfill the requirements established in section 6.1 while remaining free of charge and open source.

The next section of this chapter develops an evaluation of the software *CI Analytics* with regard to the previously defined requirements and to the actual benefits for its users.

6.3. *CI Analytics* Evaluation

This section presents an evaluation of the software *CI Analytics*. Part 6.3.1 contains an analysis of the software *CI Analytics* with regard

---

[7]   Further details are available at *http://wpf.amcharts.com/quick/* (accessed on 24.09.2011).

to the functional and non-functional requirements identified in section 6.1. Subsequently, part 6.3.2 introduces the results of an empirical survey performed to assess the businesss benefits of the software *CI Analytics*.

6.3.1. Requirements Assessment

In section 6.1, 15 functional and non-functional requirements that should be fulfilled by the software *CI Analytics* have been defined. Table 6.4 summarizes to which extent these requirements have been completed in the actual version of the software *CI Analytics*. As outlined in this table, all requirements have been at least partly achieved and nine out of the 15 requirements are fully achieved. The next paragraphs provide further details on each of these achievements:

Table 6.4.: Fulfillment of the Functional and Non-Functional Requirements

No.	Name	Fully Achieved	Partly Achieved	Not achieved
Data Source Access				
1	Access and process data stored in the application CAS genesisWorld (*functional*)	✓		
2	Support the access to additional data sources (*functional*)	✓		
3	Minimize performance impact on CAS genesisWorld (*non-functional*)	✓		
4	Provide scalable algorithm to access the data (*non-functional*)	✓		
5	Ensure that sensitive data is securely handled (*non-functional*)		✓	

Fulfillment of the Functional and Non-Functional Requirements (Continued)

No.	Name	Achievement		
		Fully Achieved	Partly Achieved	Not achieved
Customer Intimacy Calculation				
6	Consider calibration parameters to perform the metrics calculation (*functional*)	✓		
7	Calculate the acquired customer intimacy metrics at the individual level (*functional*)	✓		
8	Calculate the acquired customer intimacy at the organizational level, including the network based centrality metrics (*functional*)		✓	
9	Calculate the leveraged customer intimacy metrics (*functional*)		✓	
10	Use efficient algorithms and scalable architecture to calculate the customer intimacy metrics (*non-functional*)	✓		
11	Incrementally update the customer intimacy metrics values (*non-functional*)		✓	
Customer Intimacy Representation				
12	Visualize the acquired customer intimacy metrics at the individual level by means of a graph representation (*functional*)	✓		
13	Visualize the acquired customer intimacy metrics at the organizational level (*functional*)		✓	
14	Visualize the leveraged customer intimacy metrics by means of charts (*functional*)		✓	

Fulfillment of the Functional and Non-Functional Requirements (Continued)

No.	Name	Achievement		
		Fully Achieved	Partly Achieved	Not achieved
15	Provide data visualization by means of a web-based interface (*functional*)	✓		

1. **Access and process data stored in the application CAS genesisWorld** *(Fully Achieved)*
   The component *CI ETL* provides the ability to access all required data stored in CAS genesisWorld to calculate the customer intimacy metrics.

2. **Support the access to additional data sources** *(Fully Achieved)*
   The modular architecture of the software *CI Analytics* based on the *CI ETL* and the *CI Data Warehouse* allows to easily add new sources of data by updating the *CI ETL* component and, if required, by adding new tables in the *CI Data Warehouse*.

3. **Minimize performance impact on CAS genesisWorld** *(Fully Achieved)*
   Once the data contained in CAS genesisWorld has been retrieved by the *CI ETL* component and loaded in the *CI Data Warehouse*, the software *CI Analytics* is completely independent from CAS genesisWorld and, thus, does not impact its performance. The *CI ETL* component can be scheduled to run during time period where CAS genesisWorld is not used, for instance during the night.

4. **Provide scalable algorithm to access the data** *(Fully Achieved)*
   The access to data is based upon standard business intelligence methods, using an ETL process and a data warehouse, and it relies on established technology. It is therefore capable of efficiently handling large amount of data. In a test, 20 minutes were necessary to process data related to 14 customer organizations stored in a real CAS genesisWorld database, whereas the first prototype introduced in appendix E.3 required 20 hours to complete the same operation.

5. **Ensure that sensitive data is securely handled** *(Partly Achieved)*
   The data in the *CI Data Warehouse* is securely stored as this component is protected by the security mechanisms implemented in Microsoft SQL Server 2008. However, in the current version, the different components *CI ETL*, *CI Data Warehouse*, *CI Services* and the *CI Dashboard* do not communicate using secured protocols. In addition the *CI Dashboard* does not implement an authentication mechanism. This aspect is, however, not in the scope of this thesis.

6. **Consider calibration parameters to perform the metrics calculation** *(Fully Achieved)*
   The *CI Services* take the calibration parameters such as the time period or the time segment size as inputs and use them to calculate the customer intimacy metrics. Further details are provided in appendix E.

7. **Calculate the acquired customer intimacy metrics at the individual level** *(Fully Achieved)*
   As outlined in table 6.3, eight *CI Services* provide the functionality to calculate the acquired customer intimacy at the individual level.

8. **Calculate the acquired customer intimacy at the organizational level, including the network based centrality metrics** *(Partly Achieved)*
   Eight *CI Services* provide the ability to calculate the eight customer interaction time based acquired customer intimacy metrics at the organizational level. The network-based centrality

metrics are, however, not yet implemented in the current version of the software *CI Analytics*. To ensure the completeness of this thesis, they are implemented in the first prototype *CI Graph* presented in appendix E.3.

9. **Calculate the leveraged customer intimacy metrics *(Partly Achieved)***
   Seven *CI Services* provide the ability to calculate seven out of the eight leveraged customer intimacy metrics. The metric proactiveness ratio is not implemented in the current version of *CI Analytics* because CAS genesisWorld does not contain the relevant data for its calculation.

10. **Use efficient algorithms and scalable architecture to calculate the customer intimacy metrics *(Fully Achieved)***
    The algorithms implemented in the *CI Services* are capable of handling large amount of data and efficiently process the metrics calculation. Further details on these algorithms are available upon request from the author.

11. **Incrementally update the customer intimacy metrics values (Partly Achieved)**
    The *CI ETL* process can be scheduled to run automatically in different time intervals. The configuration of the scheduler is, however, not yet implemented in the *CI Analytics*. This could be achieved in a future version via an additional administration interface in the *CI Dashboard*.

12. **Visualize the acquired customer intimacy metrics at the individual level by means of a graph representation*(Fully Achieved)***
    As illustrated in figure 6.4, the *CI Dashboard* provides the functionality to represent the acquired customer intimacy metrics in the form of a graph representation.

13. **Visualize the acquired customer intimacy metrics at the organizational level *(Partly Achieved)***
    Figure 6.5(a) shows that by selecting a specific employee on the graph displayed by the *CI Dashboard*, a window containing additional employee related information is being displayed. This

windows can be used to represent the acquired customer intimacy metrics at the organizational level. This feature is not yet implemented in the current version of *CI Analytics* but it is, however, available in the first prototype described in appendix E.3.

14. **Visualize the leveraged customer intimacy metrics by means of charts** *(Partly Achieved)*
    The current version of the *CI Dashboard* provides the graphical representation of three out of the eight leveraged customer intimacy metrics. Its next version will include a graphical representation of the five remaining metrics.

15. **Provide data visualization by means of a web-based interface** *(Fully Achieved)*
    The *CI Dashboard* has been realized as a web-based interface using the Silverlight technology. It is, therefore, remotely accessible with any Internet browser.

6.3.2. Business Benefits Evaluation

In order to assess the potential of the software *CI Analytics*, a survey was conducted in collaboration with Thomas Herzig with 25 employees of three different IT software and services companies in July 2010. These participants were introduced to the project *CI Analytics* and were shown screenshots of the first prototype of the software *CI Analytics* which is described in appendix E.3. They were subsequently asked to evaluate the potential of this software with regard to its potential business benefits.

The assessment was performed using the questionnaire presented in appendix E.3. This questionnaire consists of 12 items which are assessed on five-point Likert-type scales.[8] These items reflect the three business benefits of the software *CI Analytics* outlined in section 1.3. 25 employees were surveyed and all filled in their questionnaires,

---

[8]   Further details on Likert-type scales are available in section 3.1.3

resulting in a 100% response rate. As detailed in appendix E.3 figure E.8, the participating employees have different roles and positions in their organization such as management, sales, services, or development. They were all involved in customer facing activities during the year preceding the survey: 88% of them were in contact with more than three customer organizations, 72% of them were in contact with more than 10 customer employees, and 64% spent over 20% of their time in customer related activities.

The three business benefits which have been considered in this survey are the following:

- **Business Benefit 1:** *CI Analytics* **helps its users to gain an overview of the relationships established with customers and customer employees.**
  In order to evaluate the potential of this first benefit, the following two items were assessed:

  - Question 5: I would use this overview to identify colleagues who have knowledge about the customer organization (strategy, process, organization, behavior, etc).

  - Question 6: I would use this overview to identify colleagues who have established relationships with customer employees.

  The results to these two questions are presented in figures 6.7 and 6.8. They confirm the relevance of the model and methodology proposed by this thesis as over 90% of the participants agree or strongly agree that they would use a software such as *CI Analytics* to identify their colleagues who have some knowledge about customers, and 80% of them agree or strongly agree that they would use it to identify colleagues who have established relationships with customer employees.

Question5: I would use this overview to identify colleagues who have knowledge about the customer organization (strategy, processes, organization, behaviour, etc.)
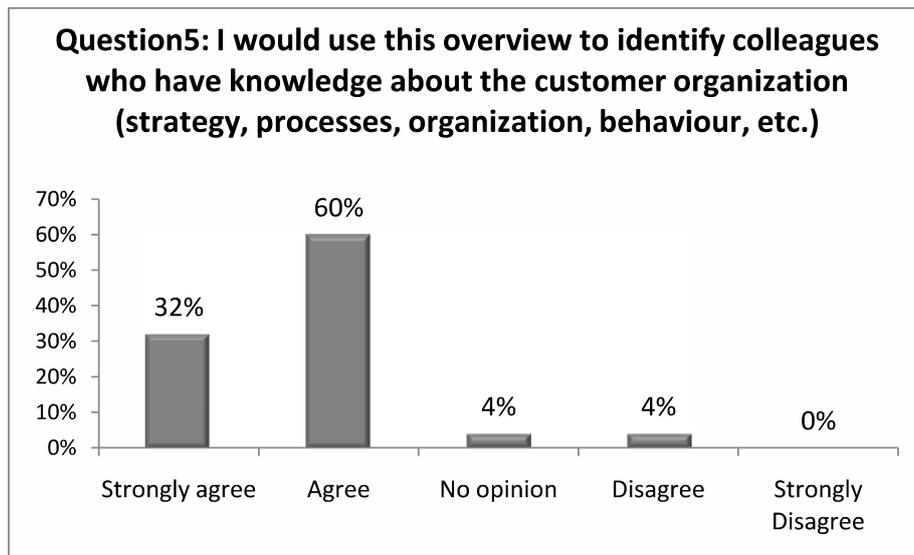
Figure 6.7.: *CI Analytics*: Business Benefit 1 – Question 5

Question6: I would use this overview to identify colleagues who have established relationships with the customer employees
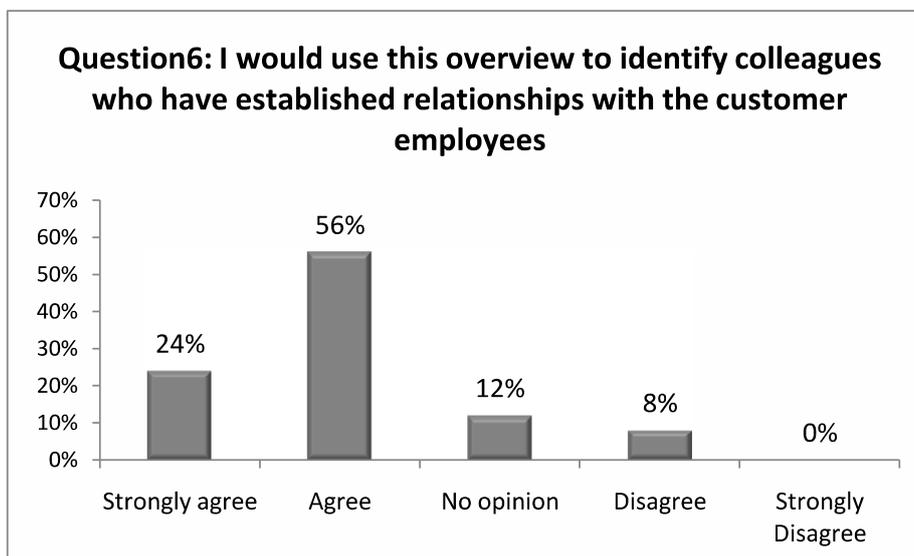
Figure 6.8.: *CI Analytics*: Business Benefit 1 – Question 6

- **Business Benefit 2:** *CI Analytics* **creates an awareness of the business relationships established by provider employees and, thus, fosters the exchange of customer related information among provider employees.**
  To evaluate this second business benefit, the following items

were assessed by the participants:

– Question 7: This relationship network overview would help us share knowledge about the customer inside our organization.

– Question 8: This relationship network overview would help us coordinate our activities towards the customer and to be seen as one team by the customer.

Figures 6.9 and 6.10 outline the results of the assessment of these two items. The results to question 7 confirm that the graph representation of the social network formed by the provider and customer employees in *CI Analytics* is a valuable knowledge management capability as 92% of the respondents agree or strongly agree that this would support the exchange of customer related knowledge in the organization. In addition, 68% of the participants also estimate that *CI Analytics* would support the coordination of the customer related activities.
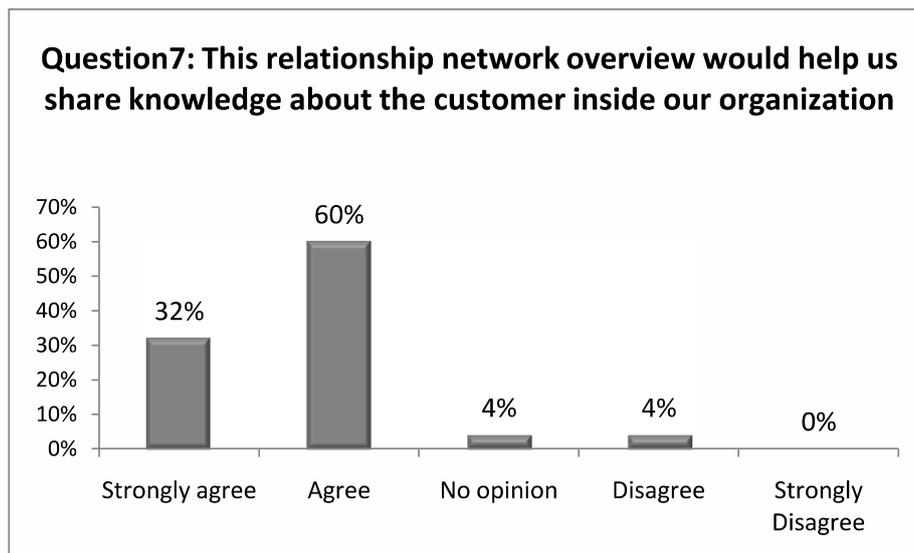
Figure 6.9.: *CI Analytics*: Business Benefit 2 – Question 7

Question8: This relationship network overview would help us coordinate our activities towards the customer and to be seen as one team by the customer
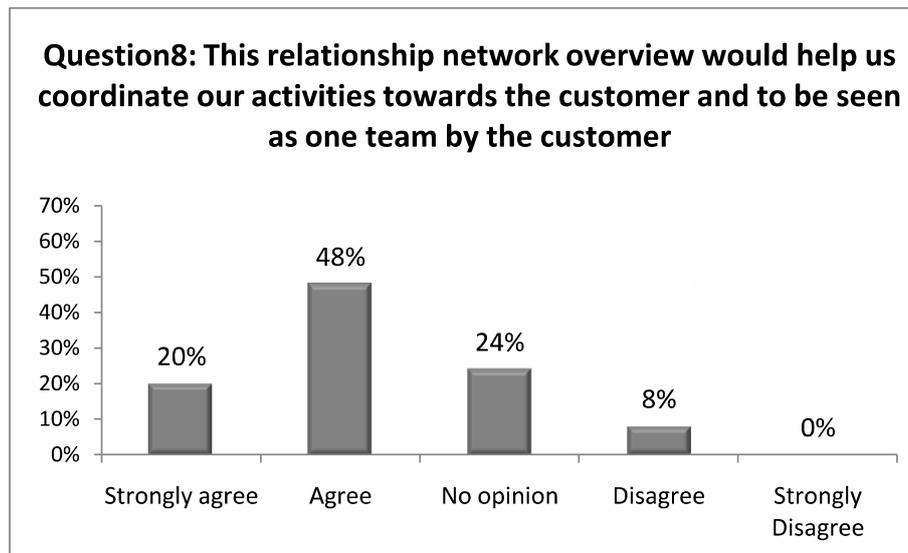
Figure 6.10.: *CI Analytics*: Business Benefit 2 – Question 8

- **Business Benefit 3:** *CI Analytics* **allows an analysis over time and a benchmarking of the relationships established with customers and supports the provider's decisions concerning investments in customers.**
  This third business benefit is evaluated with the following two items:

  – Question 9: Analyzing the evolution of this relationship network overview over time would help us monitoring the relationship with the customer.

  – Question 10: Together with other indicators such as sales results, this information would help us compare the performance achieved with different customers and would help us in our choice to invest in one or the other customer.

The results of the assessment of these two questions is outlined in figures 6.11 and 6.12. The answers to question 9 show that the survey participants overall believe in the ability of the software to monitor customer relationships, even though the results are less pronounced than for the previous items. The

answers to question 10 demonstrate that some of the respondents question the ability of the software to assess the performance achieved with different customers. This aspect may be explained by the fact that our research on the leveraged customer intimacy developed in chapter 5 was not complete at the time of the survey.
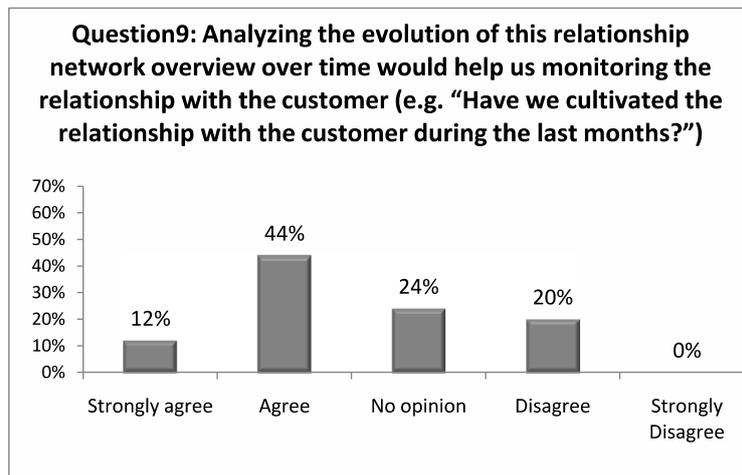
Figure 6.11.: *CI Analytics*: Business Benefit 3 – Question 9
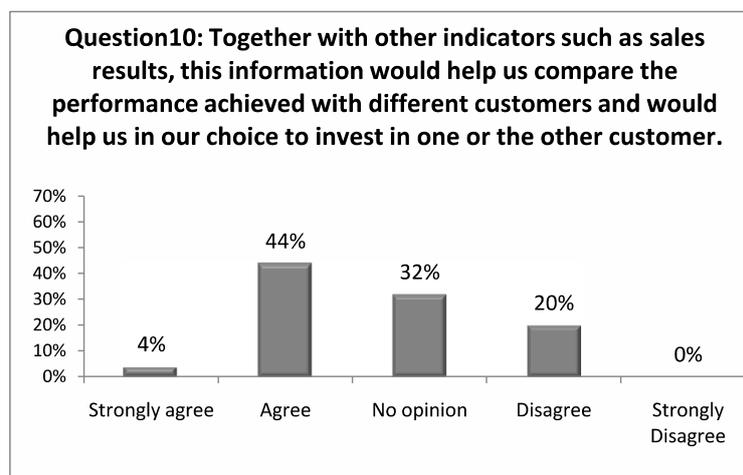
Figure 6.12.: *CI Analytics*: Business Benefit 3 – Question 10

In conclusion, the following two items were assessed by the participants in order to capture their overall appreciation of the *CI Analytics*

prototype and to determine the importance of data privacy issues related to this project:

- Question 11: I think such a visualization would be useful in our company.

- Question 12: I would have privacy concerns if this type of information was made available in my company.
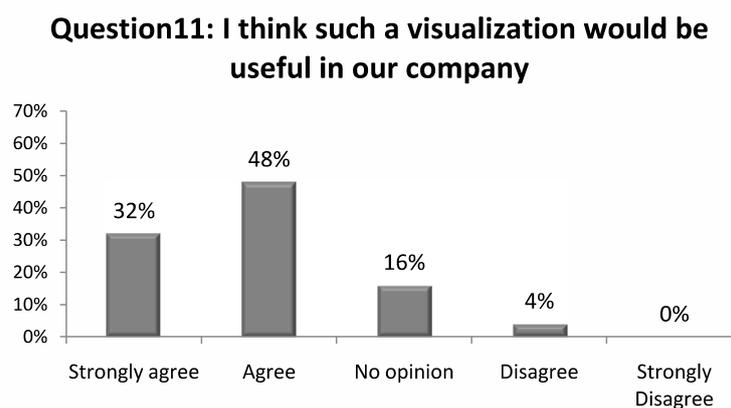
Figure 6.13.: *CI Analytics*: Overall Appreciation and Data Privacy Concerns
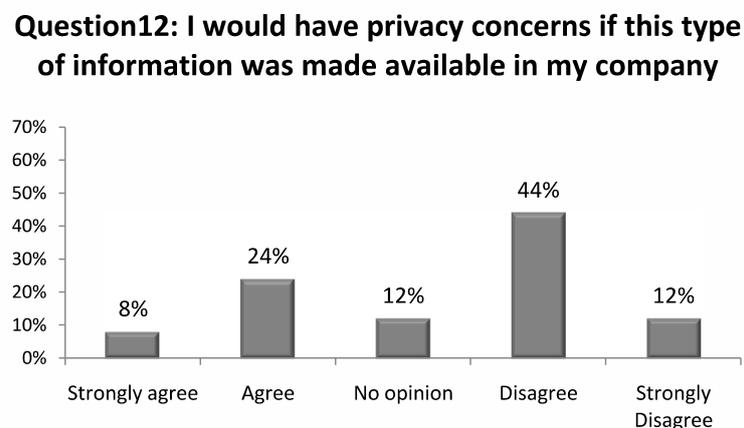
Figure 6.14.: *CI Analytics*: Overall Appreciation and Data Privacy Concerns

Figures 6.13 and 6.14 presents the results obtained for these two items. The answers to question 11 confirm the relevance of the ap-

proach proposed by this thesis as 80% of the respondents agree or strongly agree that such a visualization would be useful. In addition, the answers to question 12 demonstrate that even though data privacy issues should be thoroughly addressed by a company adopting the software *CI Analytics*, they do not seem to a a strong obstacle to the acceptance of the software by the employees as only 32% of the respondents agree or strongly agree that they would have privacy concerns if this type of information was made available in the company.

This chapter proved the feasibility of the calculation of the acquired customer intimacy metrics at the organizational and individual levels as well as of the calculation of the leveraged customer intimacy metrics through the conception and implementation of the software *CI Analytics*. Moreover, a business benefits survey confirmed that professionals involved in B2B activities would have a strong interest in such an application if it was available in their organization. The next chapter will demonstrate the relevance of the *CI Analytics* methodology proposed in chapter 5 for calibrating of the customer intimacy metrics and, thereby, accurately assessing the customer intimacy components.

7. *CI Analytics* Validation

In order to perform the assessment of the degree of customer intimacy established by a provider with his different customers in a B2B context, this thesis elaborated in chapter 5 the *CI Analytics* model and methodology. As depicted in figure 5.1, the *CI Analytics* methodology consists of seven steps. The first three steps concern the breakdown analysis of customer intimacy in multiple quantifiable components, the identification of data sources holding evidence of customer intimacy, and the determination of metrics to assess the customer intimacy components upon this data. Chapters 4 and 5 detailed the completion of these three steps and provided, thereby, the foundations for the assessment of customer intimacy.

The steps 4 to 7 of the *CI Analytics* methodology, as explained in section 5.1, support the identification of the most relevant metrics to perform an accurate inference of the customer intimacy components as well as allow a consideration of the specific activity and interaction patterns of the provider in the determination of the relative importance of the customer intimacy metrics. Step 4 refers to the actual calculation of the metrics for a specific customer. Step 5 concerns the empirical assessment of the customer intimacy components by means of a survey with provider employees. Step 6 relates to the application of machine learning algorithms on the metrics calculated in step 4 in order to predict the empirical results obtained in step 5.

Step 6, thus, results in a set of machine learning models which contain information about the most relevant metrics to accurately infer the customer intimacy components. Finally, step 7 refers to the validation and interpretation of the created machine learning models in order to derive some managerial implications. In this chapter, this sequence from step 4 to step 7 which is individually performed for each provider is referred to as the *calibration of the customer intimacy metrics*.

This chapter will show how this calibration has been performed in a real-case scenario and will evaluate to which extent the customer intimacy components *acquired knowledge of*, and *established relationships with*, customers have been inferred from the customer intimacy metrics. This chapter will, thus, validate the overall approach taken by this thesis to assessing and monitoring customer intimacy in a B2B context. This validation has been performed with the support of the IT software and services provider CAS Software AG (CAS). The customer intimacy metrics have been calculated for 14 different CAS customers upon the data stored in CAS genesisWorld. In addition, 25 CAS employees performed the empirical assessment of the customer intimacy components for these 14 customers.

The *CI Analytics* model developed in figure 5.2 establishes that the acquired knowledge of, and established relationships with, customers[1] should be assessed at two levels of detail: the individual level and the organizational level. Consequently, the calibration of the customer intimacy metrics has been performed four times in order to determine the best metrics to infer the values of these two components at these two levels of detail. Section 7.1 will elaborate on the results of the calibrations of the customer intimacy metrics performed to predict the values of the component acquired knowledge and established relationships at the individual level. Section 7.2 will subsequently develop the results of the calibrations of the customer intimacy metrics to predict these components at the organizational level.

---

[1]   These components are called acquired knowledge and established
    relationships in the remaining of this chapter.

7.1. Acquired Customer Intimacy at the Individual Level

This section presents the results of the calibration of the customer intimacy metrics to assess the acquired customer intimacy components acquired knowledge and established relationships at the individual level. Part 7.1.1 describes the data collection process which corresponds to the calculation of the customer intimacy metrics and to the empirical assessment of the customer intimacy components. Subsequently, parts 7.1.2 and 7.1.3 present the results of the application of machine learning algorithms on the calculated customer intimacy metrics to infer the values of the components acquired knowledge and established relationships.

7.1.1. Data Collection

This section consists of two parts. Part 7.1.1.1 details the setup of the calculation of the customer intimacy metrics at the individual level for 14 CAS customers. This corresponds to the step 4 of the *CI Analytics* methodology. Subsequently, part 7.1.1.2 elaborates on the survey performed to assess acquired knowledge and established relationships at the individual level for these 14 customers. This activity refers to the step 5 of the *CI Analytics* methodology.

7.1.1.1. Calculation of the Customer Intimacy Metrics

It is explained in section 5.2.2 that, given a certain set of parameters, eight metrics can be calculated upon the concepts of *customer interaction time* and *weighted customer interaction time*. These metrics are *volume, weighted volume, intensity, weighted intensity, frequency, duration, number of episodes,* and *mode of interaction*. The parameters are the time period T, the segment duration d, the default customer interaction time values of emails d_{email} and letters d_{letter}, and finally the three threshold parameters *interaction duration threshold* Δ, *interaction quantity threshold b,* and *weighted interaction quantity threshold wb*. The parameters values for calculating the customer intimacy metrics in this scenario are summarized in table 7.1. They have been chosen upon the following considerations:

Table 7.1.: Model Configurations and Metrics to Assess Acquired Customer Intimacy at the Individual Level

Configuration	A	B	C	D
Time Period T	3 months	12 months	12 months	Over one year
Segment Size d	1 month	1 month	3 months	N/A
Email CIT Value d_{email}	10 minutes	10 minutes	10 minutes	10 minutes
Letter CIT Value d_{letter}	10 minutes	10 minutes	10 minutes	10 minutes
Interaction Duration Threshold Δ	1 month	1 month	1 month	N/A
Interaction Quantity Threshold b	0	0	0	N/A
weighted Interaction Quantity Threshold wb	0	0	0	N/A
Metrics				
volume	Volume 3M	Volume 12M		Volume More 1Y
weighted volume	Volume Weighted 3M	Volume Weighted 12M		Volume Weighted More 1Y
intensity	Intensity 3M	Intensity 12M		
weighted intensity	Intensity Weighted 3M	Intensity Weighted 12M		
frequency	Frequency 3M	Frequency 12M	Frequency Quarter	
duration	Duration 3M	Duration 12M		
number of episodes	Number of Episodes 3M	Number of Episodes 12M		
mode of interaction	Mode 3M	Mode 12M		

- **Time Period** *T*
  Within the scope of this thesis, three time periods mapped to
  the operational pace of the provider organization have been
  considered:

  - the first time period is set to three months. It reflects the
    recent interactions that occurred in the past quarter and
    potentially provides the newest updates on the customer
    and his needs.

  - the second time period is set to 12 months. This time pe-
    riod corresponds to the longer projects that occurred with
    the customer during the past year.

  - the third considered time period consists of all interactions
    that occurred with the customer in the past, with the ex-
    ceptions of the interactions that happened within the last
    12 months. It is denoted as *over one year*. This time period
    reflects the fact that some employees may have established
    qualitative relationships with customer employees in the
    past, even though they had no contact within the last year.

- **Segment Size** *d*
  The segment size has to be specified in order to determine the
  level of detail of the analysis. For the 3-month and 12-month
  time periods, the main segment size is set to one month as this
  level of analysis should provide well interpretable results. In
  addition, the metric frequency is also calculated for the time
  period 12 months with a segment size of three months in order
  to gain further insights on the interaction regularity over the
  past year. With regard to the time period over one year which
  considers all interactions stored in the provider's information
  system between the beginning of the relationship with the cus-
  tomer and a year before the analysis is performed, a breakdown
  of the calculation of the *customer interaction time* and *weighted
  customer interaction time* in multiple time segments was techni-
  cally not feasible with the prototype *CI Graph*. Thus, only the
  metrics *volume* and *weighted volume* have been calculated for this
  time period.

- **Email and Letter Customer Interaction Time Values d_{email} and d_{letter}**
  The parameters d_{email} and d_{letter} provide the means to convert emails and letters exchanged with customers into customer interaction time values, thereby enabling the integration of emails and letters in the calculation of the customer intimacy metrics. In the context of this scenario, only the emails and letters exchanged with customers and containing some content which is relevant for the other provider employees are stored in the application CAS genesisWorld. Thus, both d_{email} and d_{letter} have been set in the context of this thesis to the value 10 minutes. This value reflects the average time spent by provider employees to write and read such emails and letters. Future research should investigate how to adjust the values of d_{email} and d_{letter}, taking for instance into account criteria such as the length of the email or letter, or the roles of the involved employees.

- **Threshold Parameters Δ, b, and wb**
  Finally, the different threshold parameters have to be specified. As explained in section 5.2.2.1, the *interaction duration threshold* Δ is set to one month, meaning that if no interaction occurs within one segment, a new episode starts with the next segment containing some interaction. The *interaction quantity threshold b* and *weighted interaction threshold wb* are both set to their default value 0 in order to capture and consider all interactions in the calculation of the metrics.

Table 7.1 presents an overview of the four instantiated parameter configurations as well as of the 19 resulting customer intimacy metrics. These 19 metrics have been calculated for all couples $\{p; c\}$ where p represents a CAS employee, c represents an employee of one of the 14 considered customer organizations, and existing data reveals that some interaction occurred in the past between p and c. This calculation has been performed with the software *CI Graph* presented in appendix E.3 and resulted in a data set of 10077 records. This data set is called the *customer intimacy metrics data set*. Each record in this data set consists of a reference to a CAS employee, a

reference to a customer employee, and the values of the 19 customer intimacy metrics.

7.1.1.2. Empirical Assessment

This activity corresponds to the step 5 of the *CI Analytics* methodology proposed in section 5.1.1. It refers to the empirical assessment of the customer intimacy components by means of a survey with provider employees. At the individual level, this assessment consists of an evaluation by provider employees of their knowledge of, and relationship with, customer employees. It is performed using 7-point Likert-type scales with the following four Likert items. These items are inspired from past literature and their selection is developed in section 5.2.4.[2]

- **Acquired knowledge of customer employees**

  - **Item 2.1:** "My knowledge of [CustomerEmployeeName]'s needs is thorough."

  - **Item 2.2:** "I learned a lot about [CustomerEmployeeName]'s preferences in the period I worked with him/her."

- **Established relationships with customer employees**

  - **Item 2.3:** "I have a high-quality relationship with [CustomerEmployeeName]."

  - **Item 2.4:** "I have a very collaborative relationship with [CustomerEmployeeName]."

Each provider employee participating in the empirical assessment of the acquired customer intimacy components evaluates his knowledge of, and relationship with, different customer employees using these four Likert items. Thus, each provider employees answers these four items multiple times, each time for a different customer employee. Appendix A figure A.3 illustrates such a questionnaire in which the survey participant is asked to assess his knowledge of,

---

[2]   Likert-type scales are explained in section 3.1.3.

and relationship with, seven different customer employees. Each of these empirical assessments corresponds to a couple $\{p_i; c_j\}$ where p_i represents a CAS employee and c_j represents a customer employee. Thus, as depicted in figure 7.1, each empirical assessment can be associated with a record of the customer intimacy metrics data set proposed in section 7.1.1.1 which contains the 19 calculated customer intimacy metrics.

Customer Intimacy Metrics Data Set		
p_1	c_1	Values of the 19 customer intimacy metrics for the couple (p_1, c_1)
p_i	c_j	Values of the 19 customer intimacy metrics for the couple (p_i, c_j)
...	...	...

Empirical Assessment of the Customer Intimacy Components (Survey)		
p_1	c_1	Results of the assessment by p_1 of the four Likert-items for c_1
p_i	c_j	Results of the assessment by p_i of the four Likert-items for c_j
...	...	...

Resulting Calibration Data Set			
		(prediction variables)	(underlying data of the predicted variables)
p_1	c_1	Values of the 19 customer intimacy metrics for the couple (p_1, c_1)	Results of the assessment by p_1 of the 4 Likert-items for c_1
p_i	c_j	Values of the 19 customer intimacy metrics for the couple (p_i, c_j)	Results of the assessment by p_i of the four Likert-items for c_j
...	...	...	...

Figure 7.1.: Creation of the Calibration Data Set

43 CAS employees were proposed by CAS to participate in the empirical evaluation, with the constraint that each employee performs a maximum of six assessments in order to limit their time investment. This means that, in total, a maximum of 258 assessments can be performed. Thus, only 258 out of the 10077 records in the customer intimacy metrics data set can be associated with an empirical assessment. A thorough sampling of the customer intimacy metrics data set is, therefore, necessary in order to select these 258 records.

The purposeful sampling methodology is applied in this thesis in order to manage this constraint on the sample size and select these records. Purposeful sampling refers to the "selection of information-

rich cases for study in-depth" (Patton, 2002, p.45). These information-rich cases are those which have a high relevance for the purpose of the investigation. Berry & Linoff (2004, p.63) confirm that "a smaller, balanced sample is preferable to a larger one with a very low proportion of rare outcomes." In the context of this thesis, the purpose of the analysis is the assessment of the customer intimacy components and, more specifically, the identification of the provider employees that have gathered significant knowledge of, and established good relationship with, customer employees. In the customer intimacy metrics data set, several records have very low customer intimacy metrics values, thereby indicating that very few interactions occurred between the corresponding provider and customer employees. These specific records are unlikely to be correlated with high degrees of knowledge and relationship and, therefore, should be ignored as they do not have a high relevance for our analysis. Three clusters that reflect some relevant interaction patterns between the 43 surveyed employees and the customer employees have been considered in order to create the sample:

- The first cluster contains records indicating that, over the past year, the quantity of interaction was above 2.6 hours and some face-to-face interaction occurred. These records have a high probability of being correlated to high customer intimacy values, as people have met at least once in person. The records pertaining to this cluster, therefore, fulfill the following two conditions: *Volume 12M* > 2.6 and *Mode 12M* > 0. This cluster contains in total 141 records.

- The second cluster contains records indicating that, over the past year, the quantity of interaction was above 1 hour, but no face-to-face interaction occurred. These records provide the ability to evaluate the influence of face-to-face interaction on the customer intimacy components. The records pertaining to this cluster fulfill the following two conditions: *Volume 12M* > 1 and *Mode 12M* $= 0$. This cluster contains in total 54 records.

- Finally, in order to assess the impact of the interaction that occurred before the past year, the third cluster contains the

records indicating that no interaction occurred within the last year, but the customer interaction time before the past year is above 5 hours. The records pertaining to this cluster fulfill the following two conditions: *Volume 12M* = 0 and *Volume More1Y* > 5. This cluster contains in total 73 records.

Combining the three clusters, the overall sample contains a total of 232 records. It has been used in order to generate the questionnaires of the 43 CAS employees participating in the empirical customer intimacy assessment: the CAS employees are asked to assess their knowledge of, and relationship with, provider employees which are referenced in these 232 records. Thus, each respondent receives a unique questionnaire which is tailored to his past interaction with customer employees. A custom application to generate them automatically has been implemented in order to create these questionnaires in an efficient manner.

Table 7.2.: Creation of the Calibration Data Set

Cluster	Conditions	Requested	Received
1	*Volume 12M* > 2.6 and *Mode 12M* > 0	141	50
2	*Volume 12M* > 1 and *Mode 12M* = 0	54	30
3	*Volume 12M* = 0 and *Volume More1Y* > 5	73	37
	Total	232	127

The survey was conducted between October and November 2010. 25 out of the 43 employees returned their questionnaires, resulting in 127 empirical assessments of the customer intimacy components. Table 7.2 summarizes the outcome of the survey.

As illustrated in figure 7.1, in order to perform the calibration, the results of the empirical assessment of the customer intimacy components by the provider employees are appended to the corresponding records of the customer intimacy metrics data set and, thus, associated to the 19 calculated customer intimacy metrics. The resulting *calibration data set* used for the determining the most relevant metrics

to assess the values of the customer intimacy components contains 127 records consisting of (i) the reference to a provider employee (p_i); (ii) the reference to a customer employee (c_j); (iii) the 19 corresponding customer intimacy metrics values; and (iv) the empirical assessment of the acquired customer intimacy components based on the four previously introduced Likert items.

7.1.2. Calibration: *Acquired Knowledge*

This section refers to the steps 6 and 7 of the *CI Analytics* methodology proposed in section 5.1. It describes the calibration of the customer intimacy metrics to determine the values of the component *acquired knowledge* at the individual level upon the customer intimacy metrics. This calibration consists of the application of machine learning algorithms to learn how to infer the empirically assessed acquired knowledge values, as well as the validation and interpretation of the resulting machine learning models. The pre-processing and data transformation tasks are presented in parts 7.1.2.1 and 7.1.2.2. Then, the application of machine learning algorithms as well as the validation and interpretation of the machine learning models are described in part 7.1.2.3 and part 7.1.2.4.

7.1.2.1. Preprocessing

The pre-processing activity consists of three different tasks:

- **Anonymize the Data Set**
  Since the respondents provide in their questionnaires some sensitive information about their relationships with different customer employees, the anonymity of the records in the data set has to be strictly preserved. Moreover, in the scope of this thesis, information related to the characteristics of the individual employees such as their role and position is not considered: the calibration is based exclusively on the 19 metrics. Therefore, the references to the provider and customer employees in all records of the data set are removed.

- **Manage Missing Values**
  The missing values in the context of this project are twofold:

– first, they consist of the interactions and activities that occurred between the provider and customer employees which are not recorded in the information system. The calculation of the customer intimacy metrics may be incorrect if such interaction data is missing. Since most of the interaction data is automatically stored in CAS genesisWorld, it can be assumed that this type of missing value is not significant and, thus, no specific corrective action is undertaken. This aspect, though, should be considered in the results interpretation.

– the second type of missing value refers to the Likert items which have not been empirically assessed by the survey respondents. In particular, this refers to the Likert items 2.1 and 2.2 developed in section 7.1.1.2. The respondents did not assess both items 2.1 and 2.2 in 10 out of the 127 records. These 10 records are, therefore, removed from the data set because they cannot be used for the calibration. In addition, the respondents answered only one of the two items in 16 records. The method "Imputation by Using Replacement Values" proposed by Hair *et al.* (2010, p.52) is used in order to manage these missing values: this "form of imputation involves replacing missing values with estimated values based on other information available in the sample." In this context, the value of the assessed item is used to determine the value of the missing item. For instance, if the item 2.1 is answered with the value 1 and the item 2.2 is missing, then the value of the item 2.2 is also set to the value 1. As a result, the calibration data set consists of 117 records.

- **Manage Outliers**
  Outliers are "observations with a unique combination of characteristics identifiable as distinctly different from the other observations" (Hair *et al.*, 2010, p.64). Such records may be significant as they can have a strong influence on the results of analysis. These records may be deleted, transformed, or simply kept unmodified in the data set. In this project, the third option

is chosen and the outliers are considered as any other record for three reasons. First, these outliers cannot be considered as noise. These records may represent some valid patterns of interactions, even though different from others and, thus, should be considered in the training phase of the machine learning activity. Second, the objective of this calibration is to create a machine learning model that can be used to assess the customer intimacy components out of the customer intimacy metrics. If the outliers are transformed or removed from this specific data set, the machine learning model might not be accurate when applied to other data sets where the outliers are not removed or transformed. Finally, some of the chosen machine learning algorithms presented in section 3.2.2 are "robust" or "resistant to outliers" (Tan *et al.*, 2006, p.38). For instance, the decision tree C4.5 includes a pruning option in order to limit the outliers influence on the design of the machine learning model (John, 1995, p.1).

7.1.2.2. Data Transformation

The data transformation task concerns the aggregation of the two items 2.1 and 2.2 presented in section 7.1.1.2 in order to create the required target value to apply the supervised learning approach, as explained in section 3.2.1. This transformation is performed in two steps:

- **Creation of the Summated Scale** *Knowledge*
  First, a summated scale "formed by combining several individual variables into a single composite measure" is created (Hair *et al.*, 2010, p.124). The proposed summated scale is denoted *Knowledge* and is calculated as the mean of *V(Item 2.1)* and *V(Item 2.2)* which represent the empirically assessed values of the items 2.1 and 2.2 :

$$Knowledge = \frac{V(Item\ 2.1) + V(Item\ 2.2)}{2} \tag{7.1}$$

Two requirements have to be considered in the creation of the summated scale. First, the content of the summated scale has to be conceptually valid and its components should represent the same dimension. In this project, a top-down approach has been used to create the scale, and the items of the questionnaires have been determined upon existing scales presented in past literature, as explained in chapter 4. Thus, the conceptual validity of the scale is ensured. Second, the reliability of the scale should be verified. Reliability is an "assessment of the degree of consistency between multiple measurements of a variable" (Hair *et al.*, 2010, p.125). The Crombach's Alpha test is performed on the data set in order to verify the reliability of the proposed summated scale. If this test returns a value above 0.70, the summated scale is considered as reliable (Robinson *et al.*, 1991). As presented in Appendix C figure C.1, the Crombach's Alpha test on the scale *Knowledge* returned the high value of 0.912. Thus, this scale is conceptually valid and reliable.

- *Knowledge* **Scale Binarization**
  The previously created scale *Knowledge* consists of 13 categories ordered from the value 1 to the value 7 by increments of 0.5: $\{1, 1.5, 2, ..., 6, 6.5, 7\}$. As explained in section 3.2.2, since the Likert-type scales used in the questionnaires are considered as ordinal, the scale *Knowledge* is also ordinal. Thus, the purpose of the calibration is to create a machine learning model capable of predicting the class of each record in the sample. The binarization method presented in Witten *et al.* (2011, p.315) is applied in this project on the *Knowledge* scale. This binarization method converts the 13-class classification task into multiple 2-class classification tasks. The reason for the wide adoption of this technique in data mining projects is that many machine learning algorithms perform better or even are only applicable on binary classification problems (Witten *et al.*, 2011, p.315).

  The creation of a binary classification task for each class of the *Knowledge* scale would result in 13 classification tasks. For instance, a binary variable would be set to 1 if the record belongs

to the class is 1, and to 0 otherwise. Another one would be set to 1 if the record belongs to the class 1.5, and to 0 otherwise. Such a level of detail is, however, not required in this thesis: from a business perspective, the objective is to assess whether the provider employees have no knowledge, a high knowledge or a very high knowledge of specific customer employees. Thus, two binary indicators have been created:

– *Knowledge High:* This variable is designed to identify the records indicating that a provider employee has acquired some knowledge of a customer employee. The limit to consider that a provider employee has some knowledge about a customer employee is set to the median value of the *Knowledge* scale which is equal to 4. Thus, the *Knowledge High* variable is set to 1 if the variable *Knowledge* is equal or above 4.5, and it is set to 0 otherwise:

$$Knowledge\ High = \begin{cases} 1 & \text{if } Knowledge \geq 4.5 \\ 0 & \text{otherwise} \end{cases}$$

– *Knowledge Very High:* This variable serves the identification of records indicating that a provider employee has a very high knowledge of a customer employees. It is considered in this thesis that a provider employee estimates its knowledge of a provider employee as very high if he answers the items 2.1 and 2.2 of the questionnaires with values above 6. Thus, the *Knowledge Very High* variable is set to 1 if the variable *Knowledge* is equal or above 6, and it is set to 0 otherwise:

$$Knowledge\ Very\ High = \begin{cases} 1 & \text{if } Knowledge \geq 6 \\ 0 & \text{otherwise} \end{cases}$$

As described in table 7.3, within the data set consisting of 117 records, the proportions of records in which the variables *Knowledge High* and

Table 7.3.: Proportions of *Knowledge High* and *Knowledge Very High* Records

	Quantity of Records with Value 1	Quantity of Records with Value 0	Total Quantity of Records
Knowledge High	56 (48%)	61 (52%)	117 (100%)
Knowledge Very High	35 (30%)	82 (70%)	117 (100%)

Knowledge Very High are set to the value 1 are 48% and 30%. Logically, there are fewer records in which the respondent estimated having a very high knowledge than having a high knowledge of the customer employee.

The next subsections 7.1.2.3 and 7.1.2.4 focus on the creation of machine learning models in order to predict whether the records in the sample belong to the classes *Knowledge High* and *Knowledge Very High*.

7.1.2.3. *Knowledge High* Calibration and Validation

In this section, the results of customer intimacy metrics calibration to predict the value of the variable *Knowledge High* are presented. The method "10-times 10-fold cross-validation" which is explained in section 3.2.3 is applied in order to jointly create the machine learning models and validate their performance. In addition, section 3.2.3 describes the indicators used in this project to assess the performance of the resulting machine learning models. These performance indicators are the following: success rate, precision, recall, F-measure, and kappa statistic.

These performance indicators have to be considered in the context of the project in order to be interpreted. A precision of 70% may be considered as low in a certain project and high in another one, depending of the project objectives, results implications, and quality of the data set. In order to facilitate this interpretation, three intervals

which are denoted as *good*, *fair*, and *poor* are determined for each performance indicator. All calibration results which are presented in the next sections are, thus, clustered along these three intervals. However, to ensure the completeness of this thesis, the actual key performance indicator values are also detailed for all calibration results. Table 7.4 summarizes the interval values for the five performance indicators. These intervals are determined upon on the following considerations:

- **Precision**

  In this project, precision is considered as the most important indicator. Assuming that an organization considers the adoption of the *CI Analytics* model and methodology and the deployment of the software *CI Analytics* presented in chapter 6, this organization will expect precise and reliable results. Considering the size and quality of the data set, the precision is defined as good if it is above 80%, fair if it is between 60% and 80%, and poor otherwise. A precision above 80% indicates that at least four out of five records predicted as *"Knowledge High"* are actually of class *"Knowledge High"*.

- **Recall**

  Recall is considered in this project as less important than precision because the machine learning models could easily be complemented later with additional customer intimacy metrics in order to improve the capability of the model to retrieve the records of class *"Knowledge High"*. Thus, recall is considered as good if is is above 70%, fair if it is between 50% and 70%, and poor otherwise.

- **Success Rate**

  The success rate represents the overall ability of the machine learning models to predict the class of a record, regardless of its actual class. In this project, the success rate is estimated as good if it is above 75%, fair it is between 60% and 75%, and poor otherwise.

- **F-Measure**

  The F-measure, as explained in section 3.2.3, is a combination of

the precision and recall indicators, calculated as their geometric mean. Thus, the intervals good, fair, and poor of the F-Measure are also derived from the geometric means of the recall and precision indicators. The F-Measure is considered as good if it is above 75%, fair if it is between and 55% and 75%, and poor otherwise.

- **Kappa Statistic**
  The Kappa statistics compares the success rate of the machine learning algorithm with the success rate achieved by a random prediction. It is assumed that the Kappa statistic value is considered as good if the model is at least 50% better than the random predictor, fair if it is 40% to 50% better than the random predictor, and poor otherwise.

Table 7.4.: Proposed Interpretation of the Performance Indicators

| | Interpretation | | |
	Good (%)	Fair (%)	Poor (%)
Precision	$[80 - 100]$	$[60 - 80[$	$[0 - 60[$
Recall	$[70 - 100]$	$[50 - 70[$	$[0 - 50[$
Success Rate	$[75 - 100]$	$[60 - 75[$	$[0 - 60[$
F-Measure	$[75 - 100]$	$[55 - 75[$	$[0 - 55[$
Kappa Statistic	$[50 - 100]$	$[40 - 50[$	$[0 - 40[$

The four machine learning algorithms used to perform the calibration in order to predict the variable *Knowledge High* are the following: decision tree C4.5, *k*-nearest neighbor algorithm, support vector machine algorithm, and multilayer perceptron with backpropagation neural network. These algorithms are described in section 3.2.2. The data-mining application Weka[3] is used in order to perform the calibration.

---

[3]   Further information on Weka is available at
      http://www.cs.waikato.ac.nz/ml/weka/ (accessed on 23.10.2011).

In order to optimize the performance of the different machine learning algorithms, the algorithms were not only trained and tested with their default settings, but several parameter configurations were evaluated. Table C.1 in appendix C illustrates the series of configurations considered for the optimization of the decision tree C4.5. It can be observed in this table that over 50 different configurations were tested, each of them optimizing one of the parameters. The list of parameters, their descriptions as well as the parameter values considered in this thesis are detailed in appendix B. Overall, most of the configurations presented in table C.1 led to fair or good results. The model number 40 has been selected as it presents the best combination of precision and recall values (84% and 70%). Further information on these "best results" configurations for the decision tree C4.5 and for the other three machine learning algorithms is presented in table C.2. The details on each tested configuration performed with the *k*-nearest neighbor, support vector machine, and multilayer perceptron neural network algorithms are not presented in this thesis but are available upon request from the author.

Table 7.5 presents the best results achieved with each of the four machine learning algorithms. Overall, according to the interpretation intervals proposed in table 7.4, all algorithms achieve good results to predict the value of the variable *Knowledge High*. The decision tree C4.5 and multilayer perceptron neural network obtained the best results and achieved the grade good for all five indicators. On the other hand, the *k*-nearest neighbor and the support vector machine algorithms only obtained a fair recall value of 67.0%.

The "Receiver Operational Characteristic" (ROC) curve[4] of the model created with decision tree C4.5 is illustrated in figure 7.2(b). It shows that this algorithm is very efficient in order to identify the first 72% of the true positive records as the corresponding false positive ratio remains below 10%. The size of the "area under ROC" is also high with a value of 82%.

The best model created with the decision tree C4.5 algorithm is presented in figure 7.2(a). The first criteria of the tree considers the

---

[4]   Further details on the ROC curve are provided in section 3.2.3.

Table 7.5.: Prediction of the Variable *Knowledge High*: Performance Indicator Results (g=good; f=fair; p=poor)

Model	Precision (%)	Recall (%)	Success Rate (%)	F-measure (%)	Kappa (%)
C4.5	84.0 *(g)*	70.0 *(g)*	78.4 *(g)*	75.0 *(g)*	56.0 *(g)*
k-NN	83.0 *(g)*	67.0 *(f)*	77.1 *(g)*	72.0 *(f)*	54.0 *(g)*
SVM	87.0 *(g)*	67.0 *(f)*	79.1 *(g)*	74.0 *(f)*	58.0 *(g)*
NNBP	87.0 *(g)*	71.0 *(g)*	80.2 *(g)*	76.0 *(g)*	60.0 *(g)*

metric *Frequency 12M*. This indicates that interaction regularity is significant in order to obtain a high knowledge of a customer employee. More specifically, the criteria *Frequency 12M* > 25% can be interpreted as follows: if the provider employee interacted with the customer employee in at least four different months over the past year, then he considers to have a high knowledge of the customer employee.

The second criteria of the decision tree uses the metric *Volume More1Y*. If the value of the metric *Volume More1Y* is below 1.2 hours, indicating that the provider employee interacted less than 1.2 hours with the customer employee before the past year, then the provider employee does not have a high knowledge of the customer employee. The third considered criteria of the decision tree is based on the metric *Volume Weighted 12M*. This metric reflects the *weighted customer interaction time* over the past year. The value of the metric *Volume Weighted 12M* has to be above 0.375 hour so that the provider employee considers to have a high knowledge of the customer employee. Since the *weighted* metrics take the number of participating employees to each interaction into consideration, this criteria can be achieved in multiple ways: for instance, the provider employee can have a meeting of 0.375 hour alone with the customer employee, or he can meet the customer employee together with three other persons for a duration above 1.5 hours (0.375×4).

(a) Decision Tree Representation
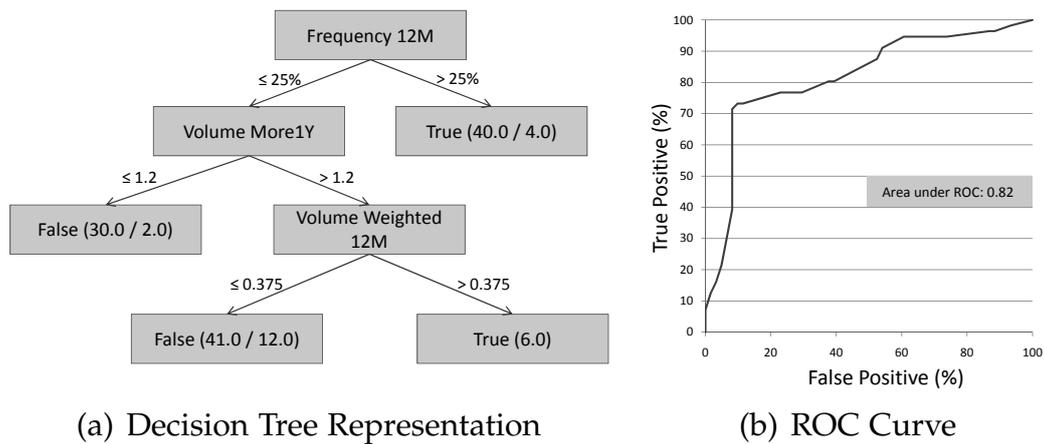
(b) ROC Curve

Figure 7.2.: *Knowledge High*: Decision Tree Model and ROC Curve

The following managerial implication can be derived from these results: a company willing to foster the acquisition of knowledge related to customer employees should encourage its own employees to regularly interact with the customer. More specifically, the provider employees should interact with the customer employees in at least four different months every year (*Frequency 12M* > 25%). These results also indicate that employees who interacted with customer employees in the past, but not within the last year still have a good knowledge of these customer employees and could be contacted if such knowledge was required in the organization.

7.1.2.4. *Knowledge Very High* Calibration and Validation

The four machine learning algorithms decision tree C4.5, *k*-nearest neighbor, support vector machine, and multilayer perceptron neural network have been trained and tested in multiple configurations in order to optimize the prediction of the variable *Knowledge Very High*. These different series of configurations are available upon request from the author. The configurations that lead to the best results with each of the algorithms are described in appendix C table C.3.

Table 7.6 summarizes the best results achieved with the four algorithms. In order to interpret these performance indicators as good, fair or poor, the intervals determined to predict the variable *Knowledge High* which are described in table 7.4 have been used. Since

the proportion of records of class *Knowledge Very High* is significantly lower than the proportion of records of class *Knowledge High* (30% vs. 48%, see table 7.3), retrieving the records of class *Knowledge Very High* is more difficult for the machine learning algorithms than retrieving the records of class *Knowledge High*. Thus, lower precision and recall values are to be expected.

While the results achieved to predict the variable *Knowledge High* were homogeneous with the four algorithms, the results achieved with regard to the prediction of the variable *Knowledge Very High* are disparate. The multilayer perceptron neural network obtained worse results than the other three algorithms according to all performance indicators. Its recall and precision values are only equal to 61.0% and 60.0%, meaning that the algorithm only retrieved 60.0% of the records of class *Knowledge Very High*, and from all records predicted as belonging to the class *Knowledge Very High*, only 61.0% of them were correct. With regard to the performance indicator success rate, the decision tree C4.5, the *k*-nearest neighbor, and support vector machine algorithm all achieved good results above 79.0%. However, the *k*-nearest neighbor is the only algorithm that achieved a good precision with a value of 83.0%, but its recall value remains only fair, with a value of 57.0%.

Table 7.6.: *Knowledge Very High*: Performance Indicator Results (g=good; f=fair; p=poor)

Model	Precision (%)	Recall (%)	Success Rate (%)	F-measure (%)	Kappa (%)
C4.5	72.0 *(f)*	55.0 *(f)*	79.4 *(g)*	59.0 *(f)*	46.0 *(f)*
k-NN	83.0 *(g)*	57.0 *(f)*	83.5 *(g)*	64.0 *(f)*	55.0 *(g)*
SVM	71.0 *(f)*	62.0 *(f)*	80.3 *(g)*	63.0 *(f)*	50.0 *(g)*
NNBP	61.0 *(f)*	60.0 *(f)*	76.5 *(g)*	58.0 *(f)*	42.0 *(f)*

The ROC curve of the best model obtained with the *k*-nearest neighbor algorithm is presented in figure 7.3(b). This model is highly

efficient in order to retrieve the first 50% of the records of class *Knowledge Very High* as the corresponding false positive ratio remains below 5%. Thus, if the objective is to identify some provider employees who have a very high knowledge of specific customer employees, this algorithm performs very well.

Figure 7.3(a) presents the decision tree created with the best configuration of the C4.5 algorithm. This model should be interpreted cautiously as it only achieves some fair results. The metric *Intensity 12M* is considered in the first node of the tree (*Intensity 12M* > 1.353). This indicates that the average interaction duration is an important aspect in order to obtain a very high knowledge of a customer employee: if the interaction of the provider employee with the customer employee last on average over than 1.353 hours, then the provider employee obtains a very high knowledge of the customer employee. If the *Intensity 12M* is below 1.353 hours, the remaining criteria of the decision tree use regularity-based metrics, thereby indicating that some regularity in the interaction is required in order for the provider employee to acquire a very high knowledge of the customer employee. The second node of the tree is based on the metric *Frequency Quarter* and tests whether the provider and the customer employee interacted in at least two different quarters over the past 12 months (*Frequency Quarter* > 25%). The third node (*Number of Episodes 3M* > 0) checks whether some interaction occurred within the last three months. Finally, the fourth node *Frequency 12M* tests whether the interaction was not too regular: if interaction occurred in more than seven months (*Frequency 12M* > 58.33%), then the provider employee does not have a very high knowledge of the customer employee. This last aspect may be interpreted as follows: the provider employees who are responsible for sending very regular information to customers, such as newsletters and advertisement do not have a very high knowledge of the provider employees.

These results lead to the following managerial implications. First, since most of the metrics used in the decision tree are based on the regularity of the interactions, these results confirm the findings of section 7.1.2.3 that a company should ensure that its employees regularly interact with customer employee in order to personally know

them. Second, since the first criteria of the tree based on the metric *Intensity 12M*, a company should organize customer events such as workshops or consulting projects in which the provider and customer employees work together on a long period of time. Such interactions allow the provider employees to obtain a very high knowledge of the customer employees.
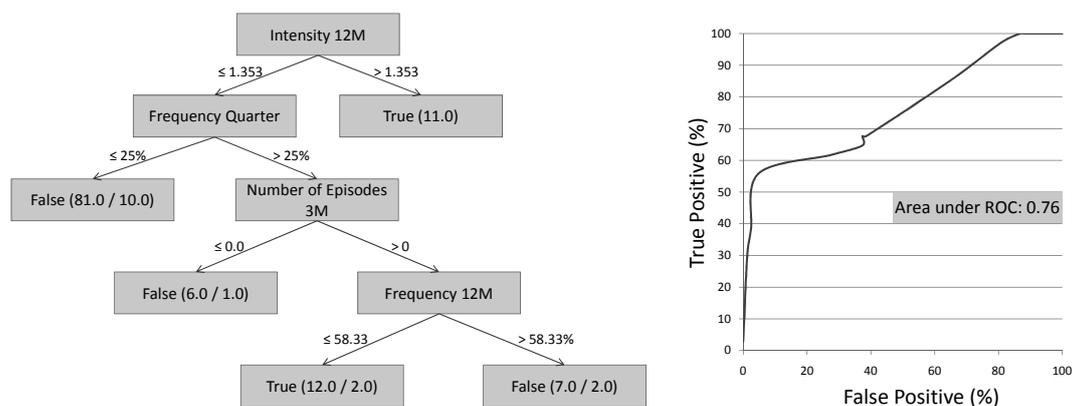
(a) Decision Tree Representation

(b) *k*-nearest Neighbor Model ROC Curve

Figure 7.3.: *Knowledge Very High*: Decision Tree Model and *k*-nearest Neighbor ROC Curve

7.1.3. Calibration: *Established Relationships*

The results of the calibration and validation of the customer intimacy metrics in order to assess the customer intimacy component *established relationships* at the individual level are presented in this section. This corresponds to the steps 6 and 7 of the *CI Analytics* methodology. Since these activities have already been thoroughly described in section 7.1.2 for the assessment of the customer intimacy component *acquired knowledge*, this section mainly focuses on the description of the results.

7.1.3.1. Preprocessing

As previously explained in section 7.1.2.1, the pre-processing activity consists of three different tasks:

- **Anonymize the Data Set**
  References to the provider and customer employees are removed from the data set in order to ensure the anonymity of the analysis.

- **Manage Missing Values**
  While no corrective action is undertaken in order to manage the missing interaction records in the database, the missing data related to the empirical assessment of the customer intimacy components is managed with the method presented in section 7.1.2.1. The items 2.3 and 2.4 have been assessed by the respondents in order to determine the value of the customer intimacy component *established relationships*. Within the original sample of 127 records, 23 records do not contain an assessment of items 2.3 and 2.4. Thus, these 23 records are removed. Then, there is no record in which either the item 2.3 or the item 2.4 has been assessed. Therefore, the final data set for the calibration of the customer intimacy component *established relationships* consists of 104 records.

- **Manage Outliers**
  The outliers are kept unchanged in the data set, as explained in section 7.1.2.1.

7.1.3.2. Data Transformation

As for the assessment of the customer intimacy component *acquired knowledge*, the data transformation activity consists of two tasks: the conception of the summated scale *Relationship* and its binarization with the creation of the indices *Relationship High* and *Relationship Very High*.

- **Creation of the summated scale *Relationship***
  The scale *Relationship* is calculated as the mean of *V(Item 2.3)* and *V(Item 2.4)* which are the empirically assessed values of the items 2.3 and 2.4. Similarly to the scale *Knowledge* presented in section 7.1.2.2, the scale *Relationship* is conceptually valid as a top down approach has been followed for its creation, and the

items 2.3 and 2.4 are derived from past literature. Moreover, this scale is also reliable as its Crombach's alpha value is equal to 0.940, as illustrated in figure C.1.

$$Relationship = \frac{V(Item\ 2.3) + V(Item\ 2.4)}{2} \qquad (7.2)$$

- ***Relationship* Scale Binarization**
  The scale *Relationship* consists of 13 classes that range from the value 1 to the value 7 by increments of 0.5: $\{1, 1.5, ..., 7\}$. The binarization method is applied in order to transform the 13-class classification task into two 2-class classification tasks. Thus, two indices, *Relationship High* and *Relationship Very High* are created:

  - *Relationship High:* The binary variable *Relationship High* distinguishes the records indicating a high quality relationship from others. Similarly to the variable *Knowledge High*, a relationship is considered as "high" if the *Relationship* value is above the median value of the Likert-scale. Thus, the variable *Relationship High* is set to 1 if *Relationship* is equal of above 4.5, and to 0 otherwise. As described in table 7.7, within the sample of 104 records, 59 records belong to the class *Relationship High*, representing 56.7% of the data set.

    $$Relationship\ High = \begin{cases} 1 & \text{if } Relationship \geq 4.5 \\ 0 & \text{otherwise} \end{cases}$$

  - *Relationship Very High:* in this project, a relationship between a provider employee and a customer employee is considered as "very high" if the value of the corresponding record on the *Relationship* scale is equal or above 6. The variable *Relationship Very High* is set to 1 if *Relationship* is equal or above 6, and to 0 otherwise. 30 records in the sample belong to the class *Relationship Very High*. They

represent a proportion of 28.8%, as illustrated in table 7.7.

$$Relationship\ Very\ High = \begin{cases} 1 & \text{if } Relationship \geq 6 \\ 0 & \text{otherwise} \end{cases}$$

Table 7.7.: Proportions of Records of Class *Relationship High* and *Relationship Very High*

	Quantity of Records with Value 1	Quantity of Records with Value 0	Total Quantity of Records
Relationship High	59 (56.7%)	45 (43.3%)	104 (100%)
Relationship Very High	30 (28.8%)	74 (71.2%)	104 (100%)

The next subsections 7.1.3.3 and 7.1.3.4 present the results of the calibration of the customer intimacy metrics in order to predict the values of the variables *Relationship High* and *Relationship Very High*.

7.1.3.3. *Relationship High* Calibration and Validation

The four machine learning algorithms decision tree C4.5, *k*-nearest neighbor, support vector machine, and multilayer perceptron with backpropagation have been trained and tested with the 10 times 10-fold crossvalidation methodology in order to calibrate the customer intimacy metrics for the prediction of the variable *Relationship High*. Several configurations were tested with each algorithm and are available upon request from the author. The best results achieved with each algorithm as well as the corresponding configurations are described in appendix C table C.4.

Table 7.8 summarizes the performance of the four algorithms to predict the variable *Relationship High*. The algorithms perform overall well, but slightly worse than to predict the variable *Knowledge High*. The decision tree C4.5, the *k*-nearest neighbor, and the support vector

machine achieve a good precision over 80.0%. The multilayer percep-
tron achieves a precision just below the good limit, with a precision
of of 79.0%. The *k*-nearest neighbor and the support vector machine
perform better than the decision tree C4.5 as they also achieve higher
recall values of respectively 75.0% and 69.0%. The *k*-nearest neigh-
bor is the only algorithm achieving both good precision and recall
values, at the cost of a fair overall success rate of 73.3%. This algo-
rithm also obtains a good F-measure value of 76.0% and a fair Kappa
statistic value of 45.0%.

Table 7.8.: *Relationship High*: Performance Indicator Results (g=good;
f=fair; p=poor)

Model	Precision (%)	Recall (%)	Success Rate (%)	F-measure (%)	Kappa (%)
C4.5	80.0 *(g)*	59.0 *(f)*	67.0 *(f)*	65.0 *(f)*	35.0 *(p)*
k-NN	80.0 *(g)*	75.0 *(g)*	73.3 *(f)*	76.0 *(g)*	45.0 *(f)*
SVM	86.0 *(g)*	69.0 *(f)*	75.4 *(g)*	75.0 *(g)*	51.0 *(g)*
NNBP	79.0 *(f)*	72.0 *(g)*	70.9 *(f)*	73.0 *(f)*	41.0 *(f)*

The ROC curve of the best *k*-nearest neighbor configuration is illus-
trated in figure 7.4(b). This diagram indicates that the algorithm is
very efficient in order to retrieve the first 50% of the records of class
Relationship High, as the corresponding false positive rate remains be-
low 10%. This algorithm, however, performs significantly worse in
order to retrieve the remaining 50% of true positive records.

Figure 7.4(a) depicts the tree created by the decision tree C4.5 algo-
rithm. Interaction regularity followed by interaction quantity are the
two main aspects leading to a qualitative relationship with customer
employees. The first node of the tree considers the metric *Frequency
Quarter*. Provider employees consider having a high quality relation-
ship with customer employees if they interacted with them in two
or more quarters over the past year (*Frequency Quarter* > 25%). If
the value of the metric *Frequency Quarter* is equal or below 25%, the
second criteria of the tree uses the metric *Number of Episodes 12M*.

(a) Decision Tree Representation

(b) *k*-nearest Neighbor Model ROC Curve
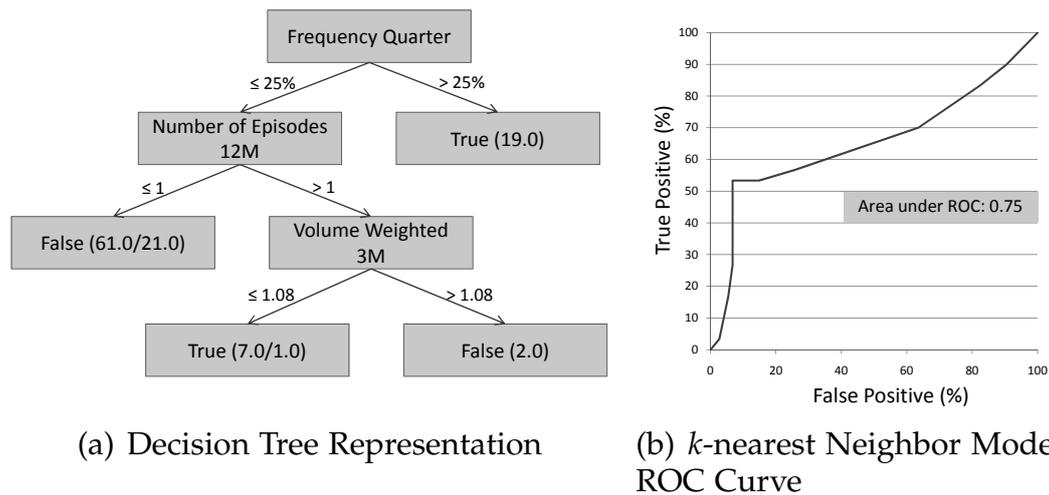
Figure 7.4.: *Relationship High*: Decision Tree Model and *k*-nearest Neighbor ROC Curve

If there was no episode or only one episode of interaction over the past year (*Number of Episodes 12M* $\leq$ 1), then the provider employees do not consider having a qualitative relationship with the customer employees. If there was more than one episode of interaction within the last year, the third criteria of the tree is based on the metric *Volume Weighted 3M*. This metric focuses on the interaction quantity over the past three months: if all previous criteria are met and if the value of the metric *Volume Weighted 3M* is below 1.08 hours, then the variable *Relationship High* is set to the value 1. This last criteria should, however, be considered cautiously as it concerns a low number of records.

From a managerial perspective, these results indicate that a regularity in the customer interaction is necessary in order for the provider employee to establish a qualitative relationship with customer employees. Since the first criteria of the tree is based on the metric *Frequency Quarter*, the regularity of the interaction over the past year is particularly important and provider employees should meet customer employees in different quarters of the year in order to develop qualitative relationships with them.

7.1.3.4. *Relationship Very High* Calibration and Validation

This section describes the results of the customer intimacy metrics calibration for the prediction of the variable *Relationship Very High*. The same four machine learning algorithms decision tree C4.5, *k*-nearest neighbor, support vector machine, and multilayer perceptron with back propagation neural network have been trained and tested by means of 10 times 10-fold cross-validation on the 104-record dataset.

Table 7.9 summarizes the best performance achieved with each algorithm. Further details on the corresponding parameter configurations are available in appendix C table C.5. As for the prediction of the variable *Knowledge Very High*, the neural network algorithm achieves poor results with a precision of 50.0% and a recall value of 51.0%. Even though the other three algorithms achieve good success rates with values comprised between 77.4% and 81.1%, none of the algorithm achieves a good precision in order to predict the value of the variable *Relationship Very High*. The decision tree C4.5 algorithm obtains the highest precision with a fair value of 75.0%. Its recall values is also fair at 52.0%. This indicates that further metrics are required in order to achieve a good performance on the prediction of the variable *Relationship Very High*.

Table 7.9.: *Relationship Very High*: Performance Indicator Results (g=good; f=fair; p=poor)

Model	Precision (%)	Recall (%)	Success Rate (%)	F-measure (%)	Kappa (%)
C4.5	75.0 *(f)*	52.0 *(f)*	81.1 *(g)*	58.0 *(f)*	48.0 *(f)*
k-NN	66.0 *(f)*	55.0 *(f)*	77.8 *(g)*	57.0 *(f)*	43.0 *(f)*
SVM	65.0 *(f)*	52.0 *(f)*	77.4 *(g)*	54.0 *(p)*	41.0 *(f)*
NNBP	50.0 *(p)*	51.0 *(f)*	74.9 *(f)*	47.0 *(p)*	34.0 *(p)*

Figure 7.5(b) illustrates the ROC curve of the best model created with the decision tree C4.5 algorithm. Similarly to the prediction of the

variable *Relationship High*, this diagram indicates that the algorithm performs very well to identify the first 52% of records of class *Relationship Very High*, but it is inefficient to retrieve the remaining ones.

(a) Decision Tree Representation   (b) Decision Tree ROC Curve
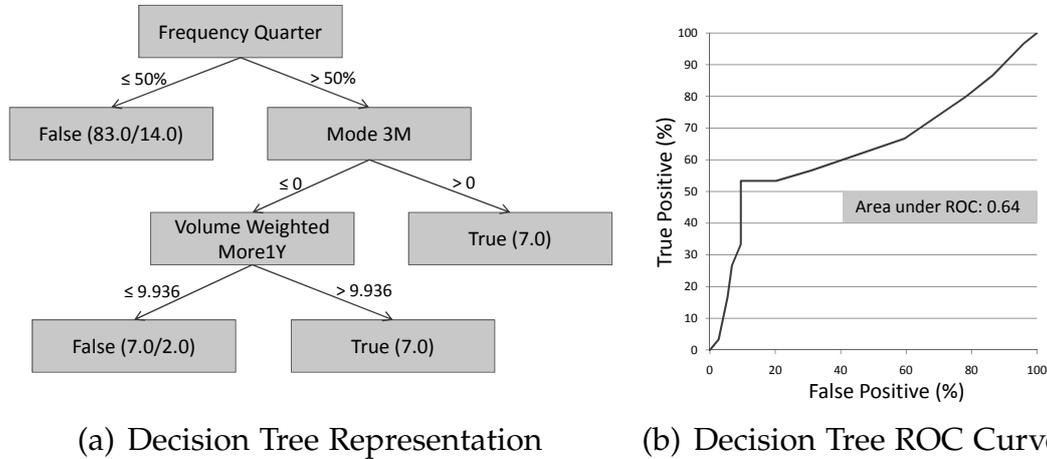
Figure 7.5.: *Relationship Very High*: Decision Tree Model and ROC Curve

Figure 7.5(a) presents the decision tree resulting from the best configuration of the C4.5 algorithm. As for the prediction of the variable *Relationship High*, the first criteria of the tree is based on the metric *Frequency Quarter*: provider employees who have established a very high relationship with customer employees interacted with them in at least three different quarters over the past year (*Frequency Quarter > 50%*). This condition is, however, not sufficient for the provider employee to consider having a very high quality relationship with the customer employee: it is also necessary that either some face-to-face interaction happened in the past three months (*Mode 3M > 0*) or that a fairly high volume of interaction occurred with the customer before the last year (*Volume Weighted More1Y > 9.936*).

These results lead to the following managerial implications: in order to develop very good relationships with customer employees, the provider should try to develop long term projects with the customer, in which provider employees have to opportunity to interact and especially meet in person with customer employees in at least three of the four quarters of the year. This confirms that a transactional approach in which the provider employee meets the customer

employee only once or twice does not allow a development of qualitative relationships.

In addition, it is possible to draw the following conclusions from the comparison of the results of the predictions of *acquired knowledge* and *established relationships* obtained in sections 7.1.2 and 7.1.3. First, the study reveals that a provider employee having a good knowledge of the customer employee has not automatically established a good relationship with this customer employee. Reciprocally, having established a qualitative relationship does not imply having a good knowledge of the customer employee. Thus, this analysis confirms the relevance of distinguishing *acquired knowledge* and *established relationships* at the individual level. Secondly, this analysis demonstrates that acquiring knowledge of a customer employee requires a different pattern of interaction than to establish a qualitative relationship with this employee. While the decision trees created to predict the variable *Knowledge High* and *Knowledge Very High* emphasize the need for frequent and intensive interactions in order to acquire customer knowledge, the decision trees created to predict the variable *Relationship High* and *Relationship Very High* use the metric *Frequency Quarter* in their first criteria, thereby highlighting the necessity for the provider employee to meet the customer in multiple quarters of the year in order to establish qualitative relationships.

The next section of this chapter develops the results of the calibration of the customer intimacy metrics to assess *acquired knowledge* and *established relationships* at the organizational level.

7.2. Acquired Customer Intimacy at the Organizational Level

While section 7.1 presents the results of the customer intimacy metrics calibration in order to predict the acquired customer intimacy components at the individual level, this section details the calibration to predict the acquired customer intimacy at the organizational level: the objective is to assess to which extent a provider employee has acquired some knowledge of, and established relationships with, a

customer organization. Following the *CI Analytics* methodology presented in section 5.1.1 and the knowledge discovery in data mining process illustrated in figure 3.2, the first part of this section focuses on the data collection task. Then, the second and third parts present the calibration results for the prediction of *acquired knowledge* and *established relationships* at the organizational level. Since these activities have already been thoroughly described in section 7.1, this section focuses on the main outcomes of the calibration and refers for details to paragraphs in section 7.1.

7.2.1. Data Collection

The data collection tasks corresponds to the steps 4 and 5 of the *CI analytics* methodology, which are the actual metric calculation and the empirical assessment of the customer intimacy components. These tasks are described in the two parts of this section.

7.2.1.1. Calculation of the Customer Intimacy Metrics

As explained in section 5.2.3, eight metrics have been designed upon the concept of *customer interaction time* in order to assess customer intimacy at the organizational level. These metrics are *volume, weighted volume, intensity, weighted intensity, frequency, duration, number of episodes and mode of interaction*. In addition, three network centrality metrics complement this list: the *degree centrality*, the *normalized degree centrality*, and the *normalized closeness centrality*. In order to perform the actual calculation, different parameters have to be determined. The four parameter configurations determined for the calculation of the metrics at the individual level and presented in section 7.1.1.1 are reused for the calculation of the metrics at the organizational level.

The centrality based customer intimacy metrics are derived from the graph representation of the customer intimacy metrics at the individual level: in order to calculate the centrality metrics, first the customer intimacy graph is created with the chosen customer intimacy metric as a weighting function. Then, the centrality metrics values are determined. In this scenario, the *degree centrality*, which reflects

the number of contacts of a provider employee in the customer organization is calculated upon the *Volume 3M* and *Volume 12M* graph representations in order to determine the number of contacts over the past 3 months and over the past year. The *normalized degree centrality* is calculated upon the *Volume 12M* graph representation. Finally, the *normalized closeness centrality* is calculated upon the *Volume 12M* and *Volume Weighted 12M* graph representations.

Table 7.10 summarizes the 24 created metrics for the assessment of the customer intimacy components. Similarly to the calculation of the customer intimacy metrics at the individual level, the customer intimacy metrics at the organizational level have been calculated for all couples $\{p, o\}$ where p represents a CAS employee, o represents one of the 14 customers of CAS, and data shows that some interactions occurred between p and some employees of o in the past. In order to perform the calculation, the prototypical application *CI Graph* which is described in appendix E.3 has been used. This calculation resulted in a data set consisting of 398 records. Each record contains a reference to a provider employee, a reference to a customer organization, and the values of the 24 customer intimacy metrics.

Table 7.10.: Model Configurations and Metrics to Assess Acquired Customer Intimacy at the Organizational Level

Configuration	A	B	C	D
Time Period T	3 Months	12 Months	12 Months	Over One Year
Segment Size d	1 Month	1 Month	3 Months	N/A
Email CIT Value d_{email}	10 minutes	10 minutes	10 minutes	10 minutes
Letter CIT Value d_{letter}	10 minutes	10 minutes	10 minutes	10 minutes
Interaction Duration Threshold Δ	1 Month	1 Month	1 Month	N/A

Model Configurations and Metrics to Assess Acquired Customer Intimacy at the Organizational Level (Continued)

Metrics	A	B	C	D
Interaction Quantity Threshold b	0	0	0	N/A
weighted Interaction Quantity Threshold wb	0	0	0	N/A

Metrics				
Volume	Volume 3M	Volume 12M		Volume More1Y
weighted Volume	Volume Weighted 3M	Volume Weighted 12M		Volume Weighted More1Y
Intensity	Intensity 3M	Intensity 12M		
weighted Intensity	Intensity Weighted 3M	Intensity Weighted 12M		
Frequency	Frequency 3M	Frequency 12M	Frequency Quarter	
Duration	Duration 3M	Duration 12M		
Number of Episodes	Number of Episodes 3M	Number of Episodes 12M		
Mode of Interaction	Mode 3M	Mode 12M		
Degree Centrality	Number of Contacts 3M (based on *Volume 3M*)	Number of Contacts 12M (based on *Volume 12M*)		

Model Configurations and Metrics to Assess Acquired Customer Intimacy at the Organizational Level (Continued)

Metrics	A	B	C	D
Normalized Degree Centrality		Degree Centrality 12M (based on *Volume 12M*)		
Normalized Closeness Centrality		Closeness Centrality 12M (based on *Volume 12M*)		
		Closeness Centrality Weighted 12M (based on *Volume Weighted 12M*)		

7.2.1.2. Empirical Assessment of the Customer Intimacy Components

The empirical assessment of the customer intimacy components refers to the step 5 of the *CI Analytics* methodology. At the organizational level, the provider employees are asked to assess with a 7-point Likert-type scale their knowledge of, and relationships with different customer organizations with the following 6 items:

- **Acquired knowledge of customer organizations**

    - **Item 1.1:** "My knowledge of [CompanyName]'s needs is thorough."

    - **Item 1.2:** "I learned a lot about [CompanyName]'s preferences in the period I worked with it."

    - **Item 1.3:** "I know the customer [CompanyName] very well."

- **Established relationships with customer organizations**

    – **Item 1.4:** "As an employee, I have a high-quality relationship with [CompanyName]."

    – **Item 1.5:** "As an employee, I have a very collaborative relationship with [CompanyName]."

    – **Item 1.6:** "I am satisfied with the relationship I have with [CompanyName]."

Further details on the item selection is presented in section 4.3. The use of Likert-type scales is motivated in section 3.1.3, and an illustrative questionnaire is presented in appendix A figure A.2.

CAS suggested 43 employees to participate to the empirical estimation, with the constraint that each employee performs a maximum of three assessments at the organizational level in order to limit the time investment. 127 records in the data set containing the calculated customer intimacy metrics at the organizational level correspond to these 43 CAS employees. These records are selected out of the 398 available records in order prepare the 43 questionnaires. The actual survey was performed between October and November 2010. 25 out of the 43 surveyed employees returned their questionnaire resulting in 77 empirical assessments. As a result, the final data set to perform the calibration of the metrics in order to assess the customer intimacy components at the organizational level consists of 77 records. Each record contains a reference to a CAS employee, a reference to one of the 14 CAS customers, the 24 calculated customer intimacy metrics, and the values of the six empirically assessed Likert items.

7.2.2. Calibration: *Acquired Knowledge*

This section presents the results of the calibration of the customer intimacy metrics in order to assess the customer intimacy component *acquired knowledge*. This corresponds to the steps 6 and 7 of the *CI Analytics* methodology. After the preprocessing and transformation tasks are explained in the first two parts, the creation of machine learning models and their validation are explained in the last two

parts of this section. Since these activities have already been thoroughly described in section 7.1.2, this section focuses on the main outcomes of the calibration.

7.2.2.1. Preprocessing

The preprocessing activity consists of three main tasks:

- **Anonymize the Data Set**
  The references to the respondents and to the customer organizations are removed from each record in the data set for the reasons outlined in section 7.1.2.1.

- **Manage Missing Values**
  As explained in section 7.1.2.1, there are two types of missing values: first, the interactions which are not recorded in the customer information system, and which may influence the calculation of the customer intimacy metrics. Similarly to the calibrations at the individual level, no action is performed in order to manage this type of missing values. Secondly, missing values refer to the Likert items which were not assessed by the respondents in the scope of the empirical evaluation of the customer intimacy components. The items 1.1, 1.2, and 1.3 were used in order to assess the component *acquired knowledge*. The data set contains only three missing values: the item 1.2 has not been evaluated in three records. Following the method "Imputation by Using Replacement Values" explained in section 7.1.2.1, the value of the item 1.2 in these three records is calculated as the average of the values of the items 1.1 and 1.3.

- **Manage Outliers**
  The outliers are kept unchanged in the data set for the three reasons explained in section 7.1.2.1.

7.2.2.2. Data Transformation

The objective of data transformation is to determine the target prediction values which are used for the calibration of the customer intimacy metrics. Similarly to the other calibrations presented in this thesis, the data transformation activity consists of two tasks:

- **Creation of the Summated Scale *Knowledge***
  The summated scale *Knowledge* is created as the mean of *V(Item 1.1)*, *V(Item 1.2)*, and *V(Item 1.3)* which are the empirically assessed values of the previously defined Likert items 1.1, 1.2, and 1.3. This scale is conceptually valid as these items have already been used to assess knowledge in past literature, and reliable as its Crombach's alpha value is equal to 0.911 as illustrated in appendix D figure D.1

$$Knowledge = \frac{V(Item\ 1.1) + V(Item\ 1.2) + V(Item\ 1.3)}{3} \quad (7.3)$$

- ***Knowledge* Scale Binarization**
  The scale *Knowledge* consists of 19 ordinal classes that range from the value 1 to the value 7 by increments of 0.33: $\{1, 1.33, ..., 6.66, 7\}$. Similarly to the data transformation applied at the individual level, the binarization method is applied in order to convert this 19-class classification task into two 2-class classification tasks with the creation of two binary variables:

  - *Knowledge High:* At the individual level, it is considered that a record belongs to the class *Knowledge High* if the value of the variable *Knowledge* is equal or above the median value of 4.5. At the organizational level, the variable *Knowledge* can take the values 4, 4.33 and 4.66 but not 4.5. Thus, the limit to distinguish the records of class *Knowledge High* at the organizational level is set to 4.66. The variable *Knowledge High* is set to 1 if the value of the variable *Knowledge* is equal or above 4.66 and to 0 otherwise. Within the calibration data set, 27 out of the 77 records belong to the class *Knowledge High*. This represents 35.1% of the data set of 77 records.

$$Knowledge\ High = \begin{cases} 1 & \text{if } Knowledge \geq 4.66 \\ 0 & \text{otherwise} \end{cases}$$

– *Knowledge Very High:* Similarly to the *Knowledge Very High* index at the individual level, a provider employee is considered as having a very high knowledge of the customer organization if its average assessment of the items 1.1, 1.2, and 1.3 is equal or above 6. Thus, the variable *Knowledge Very High* is set to 1 if *Knowledge* is equal or above 6 and to 0 otherwise. The number of records of class *Knowledge Very High* in the data set is equal to 13. This represents 16.9% of the data set.

$$Knowledge\ Very\ High = \begin{cases} 1 & \text{if } Knowledge \geq 6 \\ 0 & \text{otherwise} \end{cases}$$

Table 7.11.: Proportions of Records of Class *Knowledge High* and *Knowledge Very High*

	Quantity of Records with Value 1	Quantity of Records with Value 0	Total Quantity of Records
Knowledge High	27 (35.1%)	50 (64.9%)	77 (100%)
Knowledge Very High	13 (16.9%)	64 (83.1%)	77 (100%)

The next parts of this section present the calibration results for the prediction of the values of the variables *Knowledge High* and *Knowledge Very High*.

7.2.2.3. *Knowledge High* Calibration and Validation

In order to perform the calibration of the customer intimacy metrics, the four machine learning algorithms which are presented in section 3.2.3 have been trained and tested with the 10 times 10-fold crossvalidation method. These algorithms are the decision tree C4.5,

the *k*-nearest neighbor algorithm, the support vector machine algorithm, and the multilayer perceptron with backpropagation neural network. The following performance indicators are used to determine the calibration performance: precision, recall, success rate, F-measure, and Kappa statistic. These indicators are described in section 3.2.3. In order to facilitate the interpretation of these performance indicators values, the interpretation intervals *good*, *fair*, and *poor* defined in section 7.1.2.3 are used. Table 7.12 summarizes the ranges of these intervals. The actual values of the performance indicators are also detailed for all calibration results developed in this section in order to ensure the completeness of this thesis.

Table 7.12.: Proposed Interpretation of the Performance Indicators

	Interpretation		
	Good (%)	Fair (%)	Poor (%)
Precision	$[80 - 100]$	$[60 - 80[$	$[0 - 60[$
Recall	$[70 - 100]$	$[50 - 70[$	$[0 - 50[$
Success Rate	$[75 - 100]$	$[60 - 75[$	$[0 - 60[$
F-Measure	$[75 - 100]$	$[55 - 75[$	$[0 - 55[$
Kappa Statistic	$[50 - 100]$	$[40 - 50[$	$[0 - 40[$

In order to identify the best configurations, each algorithm has been trained and tested multiple times with different parameters. Table 7.13 presents the best results achieved with each of these algorithms and table D.1 in Appendix D provides further details on these configurations. It can be observed in table 7.13 that the decision tree C4.5, the *k*-nearest neighbor, and the support vector machine algorithm perform significantly better than the multilayer perceptron neural network, even though they only achieve fair precision values ranging from 73.0% to 75.0%. The support vector machine is clearly better than the decision tree C4.5 and the *k*-nearest neighbor algorithm as its recall values is good with a value of 81.0%, while the decision tree C4.5 and *k*-nearest neighbor algorithm only achieve re-

call values of 65.0% and 51.0%. The support vector machine also achieves a good success rate of 82.1%, a good F-measure value of 75.0% and obtains a good Kappa statistic value of 61.0%.

Table 7.13.: *Knowledge High*: Performance Indicator Results (g=good; f=fair; p=poor)

Model	Precision (%)	Recall (%)	Success Rate (%)	F-measure (%)	Kappa (%)
C4.5	73.0 (*f*)	65.0 (*f*)	79.6 (*g*)	66.0 (*f*)	53.0 (*g*)
k-NN	73.0 (*f*)	51.0 (*f*)	78.9 (*g*)	58.0 (*f*)	47.0 (*f*)
SVM	75.0 (*f*)	81.0 (*g*)	82.1 (*g*)	75.0 (*g*)	61.0 (*g*)
NNBP	50.0 (*p*)	54.0 (*f*)	68.3 (*f*)	48.0 (*p*)	29.0 (*p*)

Figure 7.6(b) presents the ROC curve obtained with the best configuration of the decision tree C4.5. This figure indicates that this algorithm performs fairly well in order to retrieve the first 60% of the records of class *Knowledge High* as the corresponding false positive rate is below 20%. This performance, however, decreases when the objective is to retrieve the remaining 40% of the records of class *Knowledge High*.

The decision tree created with the best configuraton of the C4.5 algorithm is depicted in figure 7.6(a). This tree contains two criteria. First, the tree verifies whether the value of the metric *Volume Weighted 3M*, which indicates the interaction quantity over the past three months, is above 1.48 hours. If this is the case, then the decision tree predicts that the record belongs to the class *Knowledge High*. Otherwise, the decision tree considers the value of the metric *Number of Episodes 12M*. If the last 12 months contain at least three episodes, then the variable *Knowledge High* is set to the value 1. From a managerial perspective, these results indicate that for a provider employee to obtain some knowledge of a customer organization, the key aspect is that he spends some time working with this organization (*Volume Weighted 3M > 1.48*). This confirms the results proposed in

past literature and outlining that interaction quantity is positively associated with customer knowledge (Noorderhaven & Harzing, 2009, p.2).

(a) Decision Tree Representation
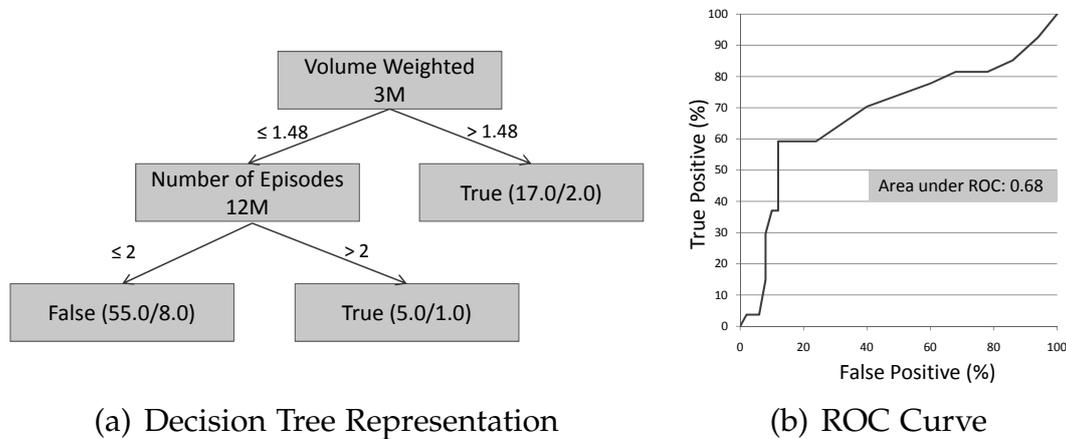
(b) ROC Curve

Figure 7.6.: *Knowledge High*: Decision Tree Model and ROC Curve

7.2.2.4. *Knowledge Very High* Calibration and Validation

Table 7.14 presents the best results achieved with the four machine learning algorithms in order to predict the value of the variable *Knowledge Very High*. Further details on the corresponding configurations are available in Appendix D table D.2. All algorithms achieve a high success rate above 80.0%. However, none of the algorithms obtains good precision and recall values. This indicates that the machine learning models are capable of predicting the value of the variable *Knowledge Very High* when this value is equal to 0, but not when this value is equal to 1. The decision tree C4.5 and the *k*-nearest neighbor algorithm obtain the highest precision with values of 41.0% and 42.0%. These values remain too low as over half of the records predicted as being of class *Knowledge Very High* are incorrectly classified. The decision tree achieves the best recall value, but this indicator remains too low with the value of 45.0%. Its Kappa statistic is also low with a value of 35.0%.

Figure 7.7(b) illustrates the ROC curve obtained with the best configuration of the decision tree C4.5 algorithm. Importantly, even though

Table 7.14.: *Knowledge Very High*:   Performance Indicator Results
           (g=good; f=fair; p=poor)

Model	Precision (%)	Recall (%)	Success Rate (%)	F-measure (%)	Kappa (%)
C4.5	41.0 *(p)*	45.0 *(p)*	84.6 *(g)*	40.0 *(p)*	35.0 *(p)*
k-NN	42.0 *(p)*	39.0 *(p)*	87.6 *(g)*	38.0 *(p)*	36.0 *(p)*
SVM	32.0 *(p)*	35.0 *(p)*	80.1 *(g)*	32.0 *(p)*	23.0 *(p)*
NNBP	14.0 *(p)*	19.0 *(p)*	80.1 *(g)*	14.0 *(p)*	10.0 *(p)*

this curve confirms the poor performance of the prediction of the
variable *Knowledge Very High*, it also indicates that the model is effec-
tive in order to retrieve the first 40.0% of the records of class *Know-
ledge Very High* since the corresponding false positive percentage is
equal to 10.0%. Thus, this algorithm can be used if the objective is to
identify a few number of employees who have acquired a very high
knowledge of a customer organization.

(a) Decision Tree Representation   (b) ROC Curve
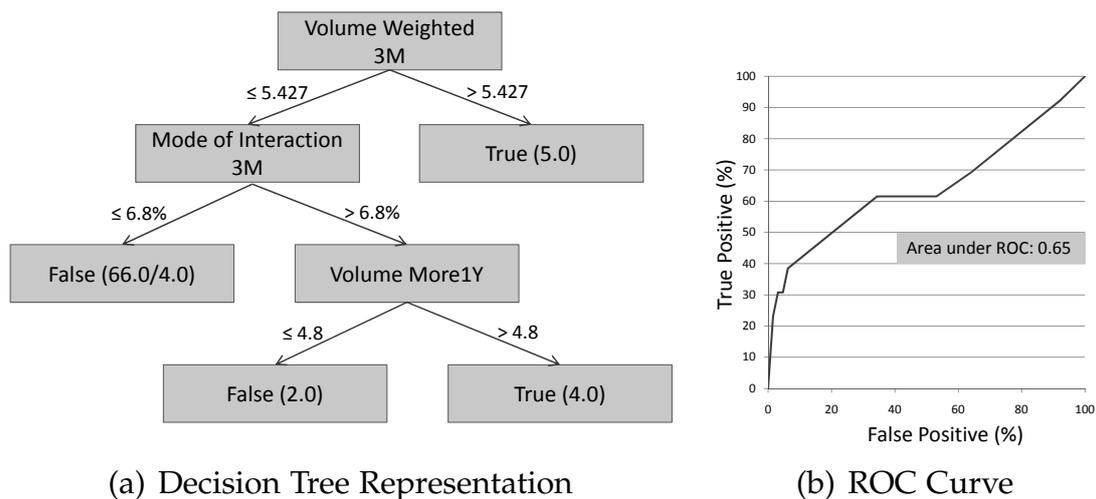
Figure 7.7.: *Knowledge Very High*:   Decision Tree Model and ROC
           Curve

As illustrated in figure 7.7(a), volume and mode of interaction are the
two main interaction characteristics used by the decision tree in order

to determine whether a record belongs to the class *Knowledge Very High*. The tree should be interpreted with caution as it did not obtain a good precision value. The first criteria of the tree uses the metric *Volume Weighted 3M*, indicating thereby that the interaction quantity within the last three month is important for a provider employee to obtain a very good knowledge of a customer organization. If the value of the metric *Volume Weighted 3M* is above 5.427 hours, then the provider employee has a very good knowledge of the customer organization. The second criteria of the tree is based on the metric *Mode of Interaction 3M*. This criteria indicates that if less than 6.68% of the interaction in the past three months occurred via face-to-face meetings (*Mode of Interaction 3M* $\leq$ 6.8%), then the variable *Knowledge Very High* is set to 0, and the provider employee does not have a very high knowledge of the customer organization. The third criteria of the decision tree considers the metric *Volume More1Y*. If the previous condition on the mode of interaction is fulfilled, the provider employee is predicted as having a very high knowledge of the customer organization if he already interacted with the customer organization for more than 4.8 hours before the past year (*Volume More1Y* $\geq$ 4.8).

From a management perspective, these results confirm the calibration results obtained for the prediction of the variable *Knowledge High*: if an organization wants to acquire some very good knowledge of its customers, it has to ensure that its employees have a high volume of interaction with the customer employees. In addition, these results show that a certain amount of face-to-face interaction is necessary for obtaining this knowledge.

7.2.3. Calibration: *Established Relationships*

This section describes the results of the calibration of the customer intimacy metrics in order to assess the customer intimacy component *established relationships* at the organizational level. While the first and second part of this section summarize the required preprocessing and data transformation tasks, the third and fourth parts present the actual results and their interpretation.

7.2.3.1. Preprocessing

Similarly to the preprocessing activity performed to assess the component *acquired knowledge* and presented in section 7.2.2.1, this preprocessing activity consists of three tasks:

- **Anonymize the Data Set**
  References to the provider employee and customer organization in each record are removed since they are not required to perform the analysis.

- **Manage Missing Values**
  As for the other preprocessing tasks presented in this chapter, no corrective action is performed in order to manage the missing interaction data in the provider's information system. The Likert items 1.4, 1.5 and 1.6 presented in section 7.2.1.2 have been empirically assessed in order determine the value of the customer intimacy component *established relationships* at the organizational level. Four out of the 77 records of the calibration data set do not contain an assessment of any of these three items and, thus, are removed from the data set. All other records contain the empirical assessment of all three items. The dataset used to perform the calibration of the customer intimacy metrics to determine the value of the component *established relationships* at the organizational level therefore consists of 73 records.

- **Manage Outliers**
  Similarly to the calibrations presented in the previous sections, the outliers are kept unchanged in the data set for the three reasons explained in section 7.1.2.1.

7.2.3.2. Data Transformation

The data transformation activity relates to the transformation of the empirical data into variables used to calibrate the customer intimacy metrics in order to assess the customer intimacy component *established relationships*. This data transformation relates to the creation

of the summated scale *Relationship* and its binarization with the creation of the variables *Relationship High* and *Relationship Very High*:

- **Creation of the Summated Scale *Relationship***
  With *V(Item 1.4)*, *V(Item 1.5)*, and *V(Item 1.6)* representing the empirically assessed values of the items 1.4, 1.5, and 1.6, the summated scale *Relationship* is calculated as the mean of these three values. The conceptual validity of this scale is ensured as the items were all already used in past literature in order to assess relationship quality. This scale is also reliable as its Crombach's alpha value is equal to 0.891 as illustrated in appendix D figure D.1.

$$Relationship \ = \ \frac{V(Item\ 1.4) + V(Item\ 1.5) + V(Item\ 1.6)}{3} \quad (7.4)$$

- ***Relationship* Scale Binarization**
  The scale *Relationship* consists of 19 ordinal classes ranging from the value 1 to the value 7 by increment of 0.33. The binarization process is performed in order to transform this 19-class classification task into two 2-class classification tasks:

  - *Relationship High:* The variable *Relationship High* is created in order to identify the provider employees which have established a high relationship with a customer organization. Similarly to the binary variable *Knowledge High*, the variable *Relationship High* is set to 1 if *Relationship* is equal or above the value 4.66 and to 0 otherwise. As presented in table 7.15, the variable *Relationship High* is set to 1 in 35 out of the 73 records of the calibration dataset, representing a proportion of 45.5% of the calibration data set.

$$Relationship\ High = \begin{cases} 1 & \text{if } Relationship \geq 4.66 \\ 0 & \text{otherwise} \end{cases}$$

  - *Relationship Very High:* The variable *Relationship Very High* distinguishes the records indicating that a provider employee has established a very high relationship with a cus-

tomer organization from the other records in the data set. It is set to 1 if the variable *Relationship* is equal or above 6 and to 0 otherwise. 16 out of the 73 records in the dataset fulfil this condition and belong to the class *Relationship Very High*. This represents a proportion of 21.9% of the calibration data set, as illustrated in table 7.15.

$$Relationship\ Very\ High = \begin{cases} 1 & \text{if } Relationship \geq 6 \\ 0 & \text{otherwise} \end{cases}$$

Table 7.15.: Proportions of *Relationship High* and *Relationship Very High* Records

	Quantity of Records with Value 1	Quantity of Records with Value 0	Total Quantity of Records
Relationship High	35 (44.5%)	38 (55.5%)	73 (100%)
Relationship Very High	16 (21.9%)	57 (78.1%)	73 (100%)

7.2.3.3. *Relationship High* Calibration and Validation

Table 7.16 presents the best results obtained with the four previously introduced algorithms. Further information on these configurations is available in appendix D table D.3. Even though multiple configurations were tested, the three algorithms decision tree C4.5, *k*-nearest neighbor, and the multilayer perceptron with backpropagation neural network all achieved a poor precision and a fair recall. They also do not obtain a good success rate and the Kappa statistic is marginal as it ranges between 11.0% and 17.0%. However, the support vector machine algorithm obtains fair to good results to predict the variable *Relationship High*. It obtains a fair precision of 64.0% and a good recall value of 74.0%. Its success rate is fair with a value of 68.8% but its Kappa statistic value remains only poor with a value of 33.0%.

Table 7.16.: *Relationship High*: Performance Indicator Results (g=good; f=fair; p=poor)

Model	Precision (%)	Recall (%)	Success Rate (%)	F-measure (%)	Kappa (%)
C4.5	55.0 *(p)*	55.0 *(f)*	56.0 *(p)*	52.0 *(p)*	11.0 *(p)*
k-NN	53.0 *(p)*	61.0 *(f)*	56.5 *(p)*	55.0 *(f)*	13.0 *(p)*
SVM	64.0 *(f)*	74.0 *(g)*	66.8 *(f)*	66.0 *(f)*	33.0 *(p)*
NNBP	54.0 *(p)*	67.0 *(f)*	58.6 *(f)*	58.0 *(f)*	17.0 *(p)*

The ROC curve of the multilayer perceptron with backpropagation neural network is presented in figure 7.8(b). This curve confirms the poor performance of the algorithm as the true positive rate is never significantly higher than the false positive rate.

(a) Decision Tree Representation (b) ROC Curve
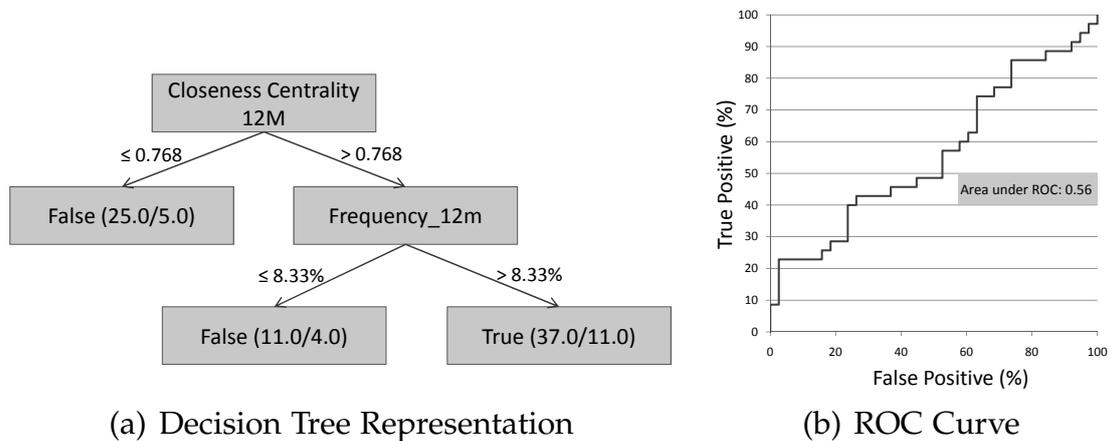
Figure 7.8.: *Relationship High*: Decision Tree Model and Multilayer Perceptron ROC Curve

Figure 7.8(a) presents the decision tree model created with the decision tree C4.5 algorithm. This model should be interpreted cautiously, as the algorithm did not achieve a particularly good performance. Importantly, the first criteria of the tree is based on the metric *Closeness Centrality 12M*. If the value of this metric is below 0.768 then *Relationship High* is set to 0. Otherwise, the decision tree considers in its second criteria the metric *Frequency 12M*. If the provider em-

ployee had interaction with employees of the customer organization in at least two different months (*Frequency 12M* $> 8.33\%$) then the variable *Relationship High* is set to 1. These results confirm the relevance of complementing the *customer interaction time* based metrics with network centrality based metrics: the topology of the social network formed by the provider and customer employees influences the perception of having established a qualitative relationship from the provider employee's perspective.

7.2.3.4. *Relationship Very High* Calibration and Validation

The four machine algorithms have been trained and tested in multiple configurations in order to predict the value of the variable *Relationship Very High* at the organizational level. The best results are presented in table 7.17 and the corresponding configurations of the algorithms are detailed in appendix D table D.4. All algorithms achieved a good success rate above 75.0%. None of them, however, achieved good precision and recall values. The best model is obtained with the decision tree C4.5 algorithm. This model has a precision of 48.0% and a recall value of 49.0%. Different reasons can explain the poor performance of this calibration. First, the current metrics are not suited for the prediction of the variable *Relationship Very High* and the model should be complemented with further metrics in order to perform the calibration. Second, the considered machine learning algorithms are not suited and other algorithms should be trained and tested. Third, the items used to assess *established relationships* at the organizational level may have been incorrectly interpreted by the participants to the survey. This leads to a wrong assessment of this component and prevents the calibration of the metrics to predict the value of the variable *Relationship Very High*.

The ROC curve of the model created with the decision tree C4.5 algorithm is presented in figure 7.9(b). This curve confirms the low performance of the algorithm. The decision tree created with the best configuration of the C.5 algorithm is depicted in figure 7.9(a). This model should be interpreted with caution as it achieved a poor performance. Regularity based metrics such as *Frequency 12M* and *Number of Episodes 12M* are not included in this tree, but *Degree Centrality*

Table 7.17.: *Relationship Very High*: Performance Indicator Results (g=good; f=fair; p=poor)

Model	Precision (%)	Recall (%)	Success Rate (%)	F-measure (%)	Kappa (%)
C4.5	48.0 *(p)*	49.0 *(p)*	78.2 *(g)*	44.0 *(p)*	33.0 *(p)*
k-NN	41.0 *(p)*	29.0 *(p)*	79.6 *(g)*	33.0 *(p)*	24.0 *(p)*
SVM	32.0 *(p)*	35.0 *(p)*	80.1 *(g)*	32.0 *(p)*	23.0 *(p)*
NNBP	18.0 *(p)*	16.0 *(p)*	77.6 *(g)*	16.0 *(p)*	11.0 *(p)*

(a) Decision Tree Representation   (b) ROC Curve
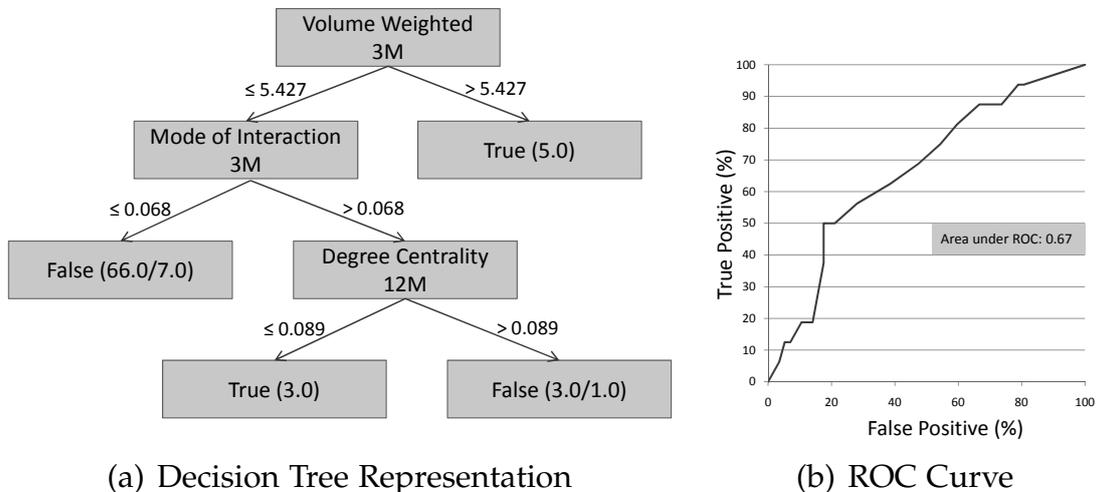
Figure 7.9.: *Relationship Very High*: Decision Tree Model and ROC Curve

12M is one of the three considered metrics. The first and second criteria of this tree use the 3-month based variables *Volume Weighted 3M* and *Mode of Interaction 3M*. Thus, the tree considers that the employees who interacted with the customer within the last three months and who had some face to face interaction are those who established very qualitative relationships with the customer organization. Importantly, this tree confirms the relevance of using network centrality based metrics in order to assess the customer intimacy components at the organizational level as the metric *Degree Centrality 12M* is used

by the tree in the third criteria.

7.3. Summary and Interpretation of the Calibration Results

Sections 7.1 and 7.2 developed the results of the customer intimacy metrics calibration for assessing the customer intimacy components *acquired knowledge* and *established relationships* at the individual and organizational levels. This section summarizes these results and further elaborates on their interpretation and managerial implications.

7.3.1. Results Summary

Table 7.18 details the best results achieved for each of the eight performed calibrations and confirms the effectiveness of the *CI Analytics* methodology to estimate the values of the customer intimacy components:

- According to the interpretation interval specified in table 7.4, three out of the eight calibrations show a good precision value above 80.0% and three of them a fair precision value comprised between 60.0% and 80.0%. Four calibrations achieve good recall values above 70.0% and two of them a fair recall value inside the 50.0% - 70.0% range.

- Six calibrations achieve a good success rate above 75.0%. The remaining two calibrations obtain a fair success rate comprised between 60.0% and 75.0%. With regard to the Kappa statistic indicator, three calibrations present good values above 50.0% for this indicator, and two calibrations a fair value in the 40.0% - 50.0% range.

- Two calibrations which concern the prediction of the variables *Knowledge Very High* and *Relationship Very High* at the organizational level lead to poor results. These calibration achieve precision and recall values comprised between 41.0% and 49.0%. Different reasons may explain this phenomenon: the sample

Table 7.18.: Summary of the Calibration Results (g=good; f=fair; p=poor)

Predicted Variable	Algorithm	Precision (%)	Recall (%)	Success Rate (%)	F-measure (%)	Kappa statistic (%)
Individual Level						
Knowledge High	NNBP	87.0 (g)	71.0 (g)	80.2 (g)	76.0 (g)	60.0 (g)
Knowledge Very High	k-NN	83.0 (g)	57.0 (f)	83.5 (g)	64.0 (f)	55.0 (g)
Relationship High	k-NN	80.0 (g)	75.0 (g)	73.3 (f)	76.0 (g)	45.0 (f)
Relationship Very High	C4.5	75.0 (f)	52.0 (f)	81.1 (g)	58.0 (f)	48.0 (f)
Organization Level						
Knowledge High	SVM	75.0 (f)	81.0 (g)	82.1 (g)	75.0 (g)	61.0 (g)
Knowledge Very High	C4.5	41.0 (p)	45.0 (p)	84.6 (g)	40.0 (p)	35.0 (p)
Relationship High	SVM	64.0 (f)	74.0 (g)	66.8 (f)	66.0 (f)	33.0 (p)
Relationship Very High	C4.5	48.0 (p)	49.0 (p)	78.2 (g)	44.0 (p)	33.0 (p)

size is too small for an effective training of the machine learning algorithms, the metrics chosen for the calibration are not suited and new metrics should be defined, or the items used for the empirical assessment were incorrectly interpreted by the respondents.

- The results are overall better for the assessment at the individual level than at the organizational level: while the four calibrations at the individual level obtain a fair or good precision, only two out of the four calibrations at the organizational level achieved a fair or good precision.

- The predictions of the variables *Knowledge High* and *Relationship*

High are better than those of the variables *Knowledge Very High* and *Relationship Very High* at both the individual and organizational levels. This may be explained by the higher number of records of type *"High"* in the dataset.

- Each of the four considered machine learning algorithms achieved the best overall results for at least one of the eight performed calibrations. The decision tree C4.5 achieved the best results three times, followed by the *k*-nearest neighbor algorithm and the support vector machine which obtained the best results twice. Finally, the multilayer perceptron with backpropagation neural network obtained the best results once, for predicting the value of the variable *Knowledge High* at the individual level.

- In order to ensure an optimized usage of the machine learning algorithms, each algorithm has been trained on average with 48 different configurations, as illustrated in appendix B table B.5. Thus, an average of 193 tests has been conducted for each predicted variable and a total of 1545 tests for the overall analysis. This aspect guarantees the completeness of the results obtained in this thesis.

Additional findings can be drawn from the analysis of the decision tree models presented in the previous sections. The number of occurrences of each metric in all decision trees is detailed in table 7.19. In this table, the metrics are sorted according to their corresponding interaction pattern as proposed in table 5.1. Even though this table does not take into account the position of the different metrics in the decision trees, the following aspects are significant:

- At the individual level, 13 out of the 19 calculated customer intimacy metrics are used in the decision trees, confirming the importance of these different metrics. Confirming past literature presented in section 5.2.2 on the impact of interaction regularity on knowledge and relationship, the regularity based metrics such as *Frequency* and *Number of Episodes* are the most important metrics as they occur seven times in the decision trees. Interaction quantity is also a significant customer intimacy indicator as the corresponding metrics occur four times

Table 7.19.: Number of Occurrences of the Metrics in the Decision Tree Models

Customer Intimacy Metric	Number of Occurrences		
	Individual Level	Organizational Level	Total
Interaction Regularity	7	2	9
Frequency Quarter	3		3
Frequency 12M	2	1	3
Number of Episodes 3M	1		1
Number of Episodes 12M	1	1	2
Interaction Quantity	4	4	8
Volume More 1Y	1	1	2
Volume Weighted 12M	1	1	2
Volume Weighted 3M	1	2	3
Volume Weighted More1Y	1		1
Mode of Interaction	1	2	3
Mode 3M	1	2	3
Intensity	1		1
Interaction Intensity 12M	1		1
Network Centrality	N/A	2	2
Closeness Centrality 12M		1	1
Degree Centrality 12M		1	1
Total	13	10	23

in the decision trees. Finally, the two other interaction patterns *Mode of Interaction* and *Interaction Intensity* are also used by the decision trees, thereby confirming their relevance for the assessment of the customer intimacy components.

- At the organizational level, the results should be interpreted with caution since the decision tree C4.5 algorithm did not

perform as well as at the individual level. 10 out of the 24 calculated metrics are used in the decision trees. However, it cannot be concluded that the remaining metrics are irrelevant since they may have been used by the other three algorithms. Contrary to the calibrations performed at the individual level, interaction quantity is the most important interaction pattern as the corresponding metrics occur four times in the decision trees. Then, the two interaction patterns *interaction regularity* and *mode of interaction* as well as the network centrality metrics are equally represented with two occurrences in the decision trees. The decision trees created at the organizational level, however, do not use metrics based on the pattern *interaction intensity*.

7.3.2. Results Interpretation

Multiple managerial implications can be drawn from the results developed in this chapter with regard to the acquisition of customer knowledge and to the establishment of customer relationships. As explained in chapter 5, the *CI Analytics* methodology is calibrated to the specific interaction patterns of the provider, which is in the context of this scenario the company CAS. Since the machine learning models created in this chapter are based on data provided by CAS, the following managerial implications are valid for CAS. Their validity for other providers should be evaluated in future research.

First, considering the acquisition of customer knowledge, this thesis shows that it is possible to thoroughly assess the degree of knowledge that a provider employee has acquired on customer employees upon the customer intimacy metrics. According to the results presented in sections 7.1.2.3 and 7.1.2.4, the provider should ensure that its employees frequently and regularly interact with the customer in order to foster the acquisition of this knowledge. The results based on the metric *Frequency 12M* show that provider employees should ideally meet the customer employees in at least four different months within a year in order to obtain a good knowledge of the customer employees. The use of the metric *Frequency Quarter* in the decision

trees shows that this knowledge is further developed if the interactions are distributed along the four quarters of the year. Moreover, the results based on the metric *Intensity 12M* show that the acquisition of knowledge about customer employees is fostered by events such as workshops or consulting projects in which provider and customer employees have the opportunity to interact for a longer duration. The provider should, therefore, try to organize such events with his customers.

Focusing on the acquisition of customer knowledge at the organizational level, the results presented in sections 7.2.2.3 and 7.2.2.4 show that it is possible to effectively identify the first 60% of the provider employees who have acquired a good knowledge of a customer organization. The results based on the metric *Volume Weighted 3M* outline that the provider should make sure that his employees have a significant amount of interaction with the customer every quarter in order to acquire a good knowledge of the organization. In addition, the use of the metric *Mode of Interaction 3M* in the decision trees confirms the importance of having a certain level of face-to-face interaction with the customer in order to obtain this knowledge. Aligned with past literature (Ballantyne, 2004; Hakansson *et al.*, 2009), these results confirm the importance of interactions for the development of customer knowledge.

With regard to the relationships established between provider and customer employees, the ROC curves depicted in sections 7.1.3.4 and 7.1.3.4 demonstrate that the customer intimacy metrics support in a very effective way the identification of the first 50% of the provider employees who have established good or very good relationships with customers. The predominance of the metrics *Frequency Quarter* and *Number of Episodes* outlines the importance of the regularity in the interactions in order to support the development of qualitative relationships between provider and customer employees. Provider employees should interact with the customer employees in at least two different quarters in one year to establish a good relationship and in at least three different quarters in one year to establish a very good relationship.

At the organizational level, the results presented in sections 7.2.3.3 show that the topology of the social network formed provider and customer employees has an important influence of the quality of the relationship established by the provider employee with the customer organization and, thus, further research should be performed in that direction. The results, however, do not allow to draw further conclusions on how to support the development of relationships between a provider employee and a customer organization.

The next chapter will conclude this thesis. It will summarize the contribution of this thesis, develop the extent to which the research questions defined in chapter 1 have been answered, and outline directions for future research.

8. Conclusion

This thesis was motivated by three main factors. First, customer intimacy has become over the past decades a prominent type of business strategy. It receives a growing interest from businesses undergoing a servitization endeavor and trying to generate competitive advantages from customer related knowledge and customer relationships. Second, a literature review developed in chapter 4 demonstrates the lack of methods, tools, and techniques enabling the assessment of customer intimacy. From the IT perspective, even though CRM systems aim to support the management of the relationship, they do not provide operational means and metrics for actually measuring the degree of customer intimacy established with different customers. Finally, the increasing importance of business analytics and social network analysis techniques in both practice and academia together with the availability of large scale database on which such analyses can be performed was the third motivating factor of this thesis. Contemporary businesses seek new solutions based on such methods and techniques to derive some knowledge out of the vast amount of gathered customer related data and to support their decision-making processes.

Combining these three factors, the central argument of this thesis is that, in a B2B context, the customer intimacy achieved by a provider organization with its different customers and its business impact can

be assessed and monitored at multiple levels of granularity – the individual and organizational levels – using social network analysis and business analytics techniques. This assessment and monitoring is achieved by leveraging customer related data available in the information system of the provider and by calculating a set of customer intimacy metrics from this data. This chapter will examine to which extent this argument has actually been validated by this thesis. Section 8.1 will revisit the three research questions defined in chapter 1 and will summarize the contribution of this thesis. Section 8.2 will subsequently elaborates on the managerial implications of this thesis. Finally, section 8.3 will address the limitations of this thesis and suggest directions for future research.

8.1. Contribution

In chapter 1, three main research questions addressing the central argument of this thesis have been defined. This section summarizes the solution proposed by this thesis to answer these questions and, thereby, outlines the contribution of this thesis.

Research Question 1 – How can the concept of customer intimacy be broken down into multiple assessable customer intimacy components?

Customer intimacy is a complex type of business strategy which aims at achieving sustainable competitive advantages by intensifying customer relationships and utilizing customer knowledge. In order to thoroughly evaluate this strategy, it is necessary to determine the provider's investments in developing his customer intimacy strategy as well as the effectiveness of this strategy with the different customers. The first research question of this thesis is therefore concerned with how customer intimacy can be broken down into multiple assessable customer intimacy components.

This question is addressed in chapter 4. Starting from the original definition of customer intimacy proposed by Treacy & Wiersema (1993, p.87) – "to tailor and shape products and services in order

to fit an increasingly fine definition of the customer" – this chapter exploits findings in customer intimacy related literature in order to determine the actual customer intimacy components. This thesis establishes that customer intimacy can be decomposed into two parts, namely the *acquired* and *leveraged* customer intimacy, and argues that both parts are required for the provider to successfully become "customer intimate" with his customers.

Acquired customer intimacy refers to obtaining this "fine definition of the customer" and consists of acquiring customer knowledge and establishing customer relationships. Customer knowledge is foundational to the development of a customer intimacy strategy because a thorough understanding of the customer is required to be able to adapt the solution provided to the customer. Customer knowledge covers multiple aspects such as the customer needs, satisfaction, expectations, strategy, and future plans. Established customer relationships is a second cornerstone of the *acquired* customer intimacy as customer intimacy is grounded in the domain of relationship marketing. Customer relationships are particularly significant in the considered context of B2B markets because they are an antecedent to customer knowledge: customer relationships allow the provider to understand his customers and, therefore, to improve his value proposition accordingly.

Leveraged customer intimacy reflects the actual benefits, competitive advantages, and means to improve the value proposition that the provider achieves by leveraging the acquired customer intimacy. It corresponds to the part "to tailor and shape products and services" of the definition of customer intimacy. The analysis performed in this thesis upon existing literature has led to the identification of six components pertaining to the leveraged customer intimacy. These components are customization, customer loyalty, proactiveness, cross-selling, customer participation, and transaction costs reduction. Using customer knowledge and customer relationships, the provider can customize his solution to the needs of the customer, increase customer loyalty, be proactive and anticipate the customer's expectations, increase revenues through cross-selling, improve his offering by involving the customer in the creation process, or reduce transac-

tional costs. The provider thereby generates a competitive advantage or improves his value proposition.

This first research question is, thus, answered by this breakdown analysis which has led to the identification of two components for the *acquired* customer intimacy and six components for the *leveraged* customer intimacy.

Research Question 2 – Which metrics can be created upon customer related data in order to infer the customer intimacy components?

The second research question of this thesis concerns the definition of metrics allowing the assessment of the customer intimacy components upon customer related data at both the individual and organizational levels. The solution to this question is elaborated in chapter 5.

The *inference* challenge developed in chapter 5 is a central issue of the assessment of the two *acquired* customer intimacy components acquired knowledge of, and established relationships with, customers. This challenge roots in the fact that no means is well recognized and established for analytically evaluating customer knowledge and customer relationships. In past literature, these components are mostly assessed in an empirical way. To circumvent this issue, this thesis proposes the *CI Analytics* model which relies on marketing literature and associates these two concepts to the four interaction characteristics quantity, intensity, regularity, and mode. Eight metrics are subsequently derived from these characteristics based on the concept of *customer interaction time* to assess acquired customer knowledge and established customer relationships. These metrics are *volume, weighted volume, intensity, weighted intensity, frequency, duration*, and *number of episodes*. At the organizational level, three additional metrics leveraging the topology of the social network formed by the provider and customer employees are defined. These metrics are the *degree centrality*, the *normalized degree centrality*, and the *normalized closeness centrality*.

In order to evaluate the *leveraged* customer intimacy components, this thesis elaborates a set of eight metrics by investigating prior research

and analyzing sources of data which are relevant for their assessment. These eight metrics are *customization revenue ratio*, *customer purchase frequency ratio*, *proactiveness ratio*, *cross-selling revenue ratio*, *cross-selling diversity ratio*, *customer participation quantity*, *customer participation ratio*, and *transaction effectiveness ratio*. The calculation of these metrics occurs upon interaction, activity, and revenue records.

To validate the feasibility of the calculation of these customer intimacy metrics and to make them available to users, the software *CI Analytics* which is detailed in chapter 6 has been conceived and implemented in the scope of this thesis. This software is built upon business intelligence applications standards, storing the relevant interaction, activity, and revenue data in a data warehouse. The software *CI Analytics* supports in its current version the calculation of the customer intimacy metrics upon the data contained in the application genesisWorld from CAS Software AG (CAS). Since the data contained in the warehouse can be updated on a regular basis at a user defined frequency, the software *CI Analytics* provides, in addition to the calculation of the customer intimacy metrics, the ability to monitor the evolution of these metrics over time.

In order to answer the second research question, this thesis, thus, establishes eight metrics to assess the acquired customer intimacy at the individual level, 11 metrics to assess acquired customer intimacy at the organizational level and eight metrics to assess the leveraged customer intimacy components. Moreover, this thesis confirms the feasibility of the assessment and monitoring of these metrics through the realization of the software *CI Analytics*.

Research Question 3 – Which combination of metrics provides the most accurate assessment of the customer intimacy components?

The third research question concerns the selection of the most relevant customer intimacy metrics and their combination in order to effectively assess the customer intimacy components. This question raises two issues which are the determination of the *relevance* of the different metrics and the *calibration* of the metrics to fit the interaction and activity patterns of each provider.

The *CI Analytics* methodology which is developed in chapter 5 is the solution proposed by this thesis to these two issues. This methodology is based on the established *knowledge discovery in database process* which outlines the required steps for analyzing data contained in databases (Fayyad *et al.*, 1996a). The *CI Analytics* methodology requires on the one hand to perform an empirical assessment of the customer intimacy components for selected customers by means of a survey with provider employees and on the other hand to calculate the customer intimacy metrics for the same customers. The two resulting data sets are subsequently merged in order to perform a supervised data-mining analysis. In this analysis, the calculated metrics are the prediction variables and the results of the empirical assessment are transformed into the predicted variables. Several machine learning algorithms are trained to predict the empirically assessed values of the customer intimacy components upon the calculated customer intimacy metrics. The resulting models are finally tested to ensure that they can be successfully applied to other data from the same provider, and interpreted in order to understand the most relevant metrics and to derive managerial implications.

The *CI Analytics* methodology has been validated in a real-case scenario with the company CAS. The results are detailed in chapter 7. The components acquired customer knowledge and established customer relationships related to 14 different customers were assessed by CAS employees and the corresponding customer intimacy metrics were calculated with the software *CI Graph* outlined in appendix E.3. Four algorithms have been trained to predict the empirically assessed customer intimacy components upon the calculated customer intimacy metrics: the decision tree C4.5, the multilayer perceptron with back propagation neural network, the *k*-nearest neighbor algorithm, and the support vector machine algorithm. The results have been evaluated using the 10-fold cross-validation technique with the performance indicators precision, recall, success rate, F-measure, and Kappa statistic.

The results developed in chapter 7 show that eight calibrations of the customer intimacy metrics have been performed to predict the acquired customer intimacy components at the individual and orga-

nizational levels. Six calibrations achieve a good success rate. Six calibrations achieve good or fair precision and recall values. Overall, the results of the calibration at the individual level are better than those at the organizational level. The four machine learning algorithms performed differently but none of them was significantly better than the others. The interaction metrics based on regularity such as *frequency* and *number of episodes* are the most relevant ones for assessing the customer intimacy components at the individual level. At the organization level, the interaction quantity metrics such as *volume* and *weighted volume* are the most significant ones.

This analysis confirms the effectiveness of the *CI Analytics* methodology for determining the best combination of metrics to assess the acquired customer intimacy components. While the implementation of the software *CI Analytics* proves the feasibility of calculating, monitoring, and representing the proposed customer intimacy metrics, the quantitative results validate the central argument of this thesis and demonstrate that it is possible to accurately assess customer intimacy at multiple level of details in an analytical manner.

The next section of this chapter elaborates on the managerial implications of the results obtained in this thesis.

8.2. Managerial Implications

The results achieved in the course of this thesis may have in the future significant managerial implications as they allow an organization pursuing a customer intimacy strategy to obtain new insights in the actual development and implementation of this strategy with its customers.

First, the software *CI Analytics* conceived in this thesis and described in chapter 6 can be used by a provider in order to assess the degree of customer intimacy established with its different customers at different levels of details and, thus, to support the future investments and business decisions. As illustrated in figure 6.4, this software allows on one side to assess the investments performed by the provider

employees in order to acquire knowledge of, and establish relationships with, the customer, and on the other side, using the leveraged customer intimacy indicators, to assess the business impact of this knowledge and of these relationships. In a best-case scenario, as outlined in figure 4.1, a provider pursuing a customer intimacy strategy should see in the *CI Analytics* dashboard high values with regard to acquired knowledge and established relationships as well as high values with regard to the leveraged customer intimacy metrics, thereby indicating that the provider effectively used his knowledge of, and relationships with, customers in order to derive competitive advantages and to improve its value proposition. However, if the *CI Analytics* dashboard indicates high knowledge and relationships values but low leveraged customer intimacy values, then the customer intimacy strategy is not effective as no or few competitive advantages are derived from the acquired knowledge and established relationships. In such cases, the provider should analyze whether the customer intimacy strategy is appropriate with the customer as some customers are not responsive to a customer intimacy strategy and are not ready to pay a premium for a tailored solution. The provider should also analyze whether the customer intimacy strategy was correctly implemented with this specific customer. It is indeed possible that an appropriate solution was not suggested to the customer, leading to low leveraged customer intimacy values. Since the metrics can be used in order to determine the customers with whom the customer intimacy strategy was most effective, this approach, in addition, allows a ranking and benchmarking of the different customers, thereby supporting the provider with regard to its future customer investments.

The second type of managerial implications relates to an improved coordination of the customer facing activities of the provider employees and a better sharing of customer knowledge inside the provider organization. By making the values of the customer intimacy metrics available inside the provider organization, for instance in the form proposed by the *CI Analytics* dashboard, the provider employees can easily identify colleagues who have acquired knowledge of, and established relationships with, the customer as well as those who were in contact with specific customer employees within a spe-

cific time frame, such as the past week or the past month. Using this information, the provider employees whose activities are related to a specific customer can find each other, exchange their knowledge, and coordinate their activities. For instance, if a provider employee p_1 has planned a meeting with a customer employee c and notices in the *CI Analytics* dashboard that another provider employee p_2 had a conversation with c in the past week, p_1 can contact p_2 to obtain the most recent information on c and use this knowledge when he meets c, thereby optimizing the interaction flow with the customer employee c.

Finally, the approach proposed by this thesis allows an organization to gain insights on how to best establish, maintain, and enhance customer relationships as well how to effectively acquire customer knowledge by optimizing the customer interactions and activities. The results presented in section 7.3.2 shows that the company CAS whose data was used to apply the *CI Analytics* methodology should focus on specific interaction patterns in order to acquire customer knowledge and establish customer relationships. For instance, CAS employees willing to acquire a good knowledge of customer employees should interact with them in at least four different months within a year. Moreover, in order to obtain a very good knowledge, they should organize events of longer durations. To establish qualitative relationships, a focus should be given to the regularity of the interaction: the provider employees should interact in three different quarters of the year with customer employees in order to establish very good relationships. These conditions are naturally not sufficient for acquiring knowledge and establishing relationships. An employee interacting in three different quarters does not always have a very good relationship with customer employees, but the probability that he does are higher if he follows these interaction patterns.

The next section of this chapter outlines the limitations of this contribution and suggests some directions for future research.

8.3. Outlook on Future Research

This thesis demonstrates that customer intimacy can be assessed and monitored at multiple levels of details in a B2B context using business analytics and social network analysis methods. It is also laying the foundations for further research investigating customer intimacy, relationships, and business performance in an analytical way. This section develops the limitations of the current approach and elaborates on future paths of research which could be followed upon this thesis. Seven main aspects have been identified.

- **Use different data sources to calculate the customer intimacy metrics**
  The software *CI Analytics* which has been conceived and implemented in the scope of this thesis is able to process data contained in the application CAS genesisWorld. A key benefit of CAS genesisWorld is that the relevant customer interaction, activity, and revenue data is stored in one single database with appropriate references to customers and customer employees. However, because the software *CI Analytics* only focuses in its current version on data contained in this database, the proposed *CI Analytics* methodology has not been applied to, and tested with, other sources of data. Future research should, therefore, concentrate on the integration of new sources of data in the proposed approach to assess customer intimacy and the next version of the software *CI Analytics* should support the access and processing of data contained in additional data sources such as CRM software, groupware, and project databases. This task is facilitated by the current architecture of the software *CI Analytics* which allows an easy integration of different sources of data.

- **Develop additional customer intimacy metrics to improve the assessment of the customer intimacy components**
  As developed in chapter 7, most of the performed calibrations to assess the customer intimacy components achieved good or fair results. However, some of these calibrations did not obtain acceptable results with regard to the five defined performance

indicators. For instance, the precision and recall values related to the prediction of the variable *Relationship Very High* at the organizational level only obtained poor results even though the corresponding success rate is good. Future research should, therefore, investigate the creation of new metrics to complement the existing ones and to improve the quality of the performed customer intimacy assessment. In particular, activity based metrics focusing on the time spent by customer employees on customer projects should be developed. Such data may easily be retrieved from project databases. In addition, different calibration parameters such as the time period T, the segment size d, or the interaction duration threshold Δ have been proposed by this thesis in order to configure the calculation of the customer intimacy metrics. In this thesis, as detailed in chapter 7, four different configurations of these parameters have been considered at both the individual and organizational levels. Future research should test additional configuration as well as further investigate the impact of these parameters on the accuracy of the metrics to assess the customer intimacy components.

- **Perform longitudinal analysis and add complex event processing**
  The software *CI Analytics* provides the means to calculate the customer intimacy metrics at regular time intervals, thereby enabling the monitoring of the proposed customer intimacy components. The validation performed in the scope of this thesis and elaborated in chapter 7, however, only considers a specific point in time in order to calculate the metrics. Future research should consequently focus on a longitudinal analysis of the customer intimacy metrics and evaluate which knowledge can be derived from this time driven analysis. This analysis would, for instance, uncover correlations between the evolution of the interaction and activity based metrics and business results. Such research could subsequently be combined with complex event processing in order to identify specific patterns among interaction and activity events which impact business activities (Et-

zion & Niblett, 2010). For instance, a change of the interaction regularity combined with a drop of the activity volumes could indicate some issues with the customer which should be proactively managed by the provider.

- **Investigate the correlation between the acquired and leveraged customer intimacy components and conceive a recommender system based on successful interaction and activity patterns**
  This thesis establishes a model to decompose customer intimacy into multiple components and develops multiple metrics enabling the assessment of these components upon interaction, activity, and revenue data. However, the analysis of correlations among the different customer intimacy components was out of the scope of this thesis. Future research focusing on these correlations is an important research topic potentially having significant managerial implications.

  First, focusing on the acquired customer intimacy, an investigation of the correlation between the acquired knowledge of, and the established relationships with, customers would provide an understanding of the influence of customer relationships on acquired customer knowledge. Second, focusing on the causal relationship between the acquired and leveraged customer intimacy, this investigation would provide insights on which degrees of customer knowledge and customer relationships are required in order to reach the benefits elaborated in the leveraged customer intimacy components. This analysis can be performed analytically rather than empirically using the proposed customer intimacy metrics. It would, therefore, provide a unique contribution by associating some specific interaction and activity patterns, such as the regularity or the volume of interactions to critical business impact factors such as cross-selling revenues, customer loyalty, or transaction costs reduction.

  These patterns could subsequently be implemented into a recommender systems which supports the determination of the

customer related activities of the provider. For instance, if the analysis establishes that a specific frequency of interaction has an impact on customer loyalty, the system could remind the corresponding provider employees to contact the customer employees at this frequency. If a specific incentive has been identified as particularly successful for facilitating opportunity closure and for reducing transaction costs, this incentive could be suggested to other provider employees which are in similar situations with their customers.

- **Elaborate a recommender system for optimizing the team in charge of a customer, for allocating provider employees to customer projects, and for coordinating the activities of these employees**
  The approach developed in this thesis provides the means to assess and monitor the degree of customer intimacy established with different customers. It also supports the exchange of customer related knowledge through the visualization of the social network formed by the provider and customer employees upon their interactions and joint activities.

  This approach could be further extended in future research by conceiving a recommender system which suggests a set of provider employees which are most likely to fit with the customer organization upon the customer intimacy metrics measured at the individual and organizational levels. Considering a specific customer, this recommender system could consider as inputs the roles and positions of the provider and customer employees, the current values of the acquired and leveraged customer intimacy metrics, and the objectives set by the provider for this customer. In return, this recommender system could provide a set of employees which have the adequate skills as well as the appropriate relationships and customer knowledge in order to effectively and successfully perform the customer project. This system would therefore support the optimization of the teams in charge of specific customers and the allocation of the provider employees to the different customer projects.

Moreover, since the customer intimacy metrics are monitored and updated at frequent intervals, this recommender system can easily gather details on the most recent interactions and activities that occurred with the customer employees. This system could therefore use this information in order to make recommendations to the provider employees before they contact customer employees, thereby supporting the coordination of the customer-facing activities. For instance, if a provider employee recently worked with several customer employees, the other provider employees should contact him prior to contacting this customer employees as he may have some valuable information and knows the details of the communication with the customer. This could be automatically supported by this recommender system.

- **Evaluate the legal aspects of the customer intimacy assessment**
  A critical aspect of the assessment and monitoring of the degree of customer intimacy resides in the use of personal interaction records such as emails or details on meetings. Under German law, this data does not belong to the provider organization but to the provider and customer employees involved in the corresponding interactions who for instance send and receive the emails. The provider organization is, thus, not directly allowed to use this data in order to perform the customer intimacy assessment. This problem is solved in this thesis through the exclusive use of data stored in the application CAS genesisWorld. Provider employees can freely decide for each interaction record whether they want it to be transferred to CAS genesisWorld. If the record is transferred to CAS genesisWorld, it is then considered as a business information and can be used by the organization. However, in order to access data contained in other sources of data such as email servers, a legal solution should be found. Thus, further research should further investigate from a legal perspective how to enable the calculation and utilization of the customer intimacy metrics in the provider organization.

- **Extend the proposed model towards B2C and C2C businesses**
  The model proposed by this thesis focuses on B2B organizations and takes into account the specific constraints of B2B businesses, such as the fact that users and purchasers of the provided solutions are different individuals in the customer organization. However, considering the size of B2C markets and the increasing importance of B2C and C2C services in mature economies, future research should focus on the extension of this approach towards B2C and C2C businesses and the development of B2C and C2C specific customer intimacy metrics. This approach could subsequently be integrated in Internet based social network applications such as LinkedIn, Facebook, or Xing.

Following an interdisciplinary approach, this thesis proposes a novel means for the assessment and monitoring of customer intimacy, combining a strategy and marketing concept with business analytics, network analysis, and software engineering. The outlook on future research developed in section 8.2 demonstrates the significance of the managerial implications of this approach and shows that this thesis lays the foundation for a wide variety of new research topics and for a new way to approach the assessment and implementation of business strategies.

Bibliography

Abraham, Gerard A. 2006. *Profit Power of Customer Intimacy: Deliver Top Line Revenue and Earnings Growth*. http://www.leader-values.com/Content/detail.asp?ContentDetailID=1125, accessed on 20/11/2011.

Adler, Lee. 1966. Symbiotic marketing. *Harvard Business Review*, **45**(2), 59–71.

Akura, M. Tolga, & Srinivasan, Kannan. 2005. Customer Intimacy and Cross-Selling Strategy. *Management Science*, **51**(6), 1007–1012.

Ali, Shawkat, & Smith, Kate A. 2006. On learning algorithm selection for classification. *Applied Soft Computing*, **6**(2), 119–138.

Alpaydin, Ethem. 2010. *Introduction to Machine Learning (Adaptive Computation and Machine Learning series)*. The MIT Press.

Anderson, Erin, & Weitz, Barton A. 1992. The Use of Pledges to Build and Sustain Commitment in Distribution Channels. *Journal of Marketing Research*, **24**(February), 18–34.

Arndt, Johan. 1979. Towards a Concept of Domesticated Markets. *Journal of Marketing*, **43**(4), 69–75.

Asratian, Armen S., Denley, Tristan M. J., & Häggkvist, Roland. 1998. *Bipartite graphs and their applications*. Cambridge University Press.

Ballantyne, David. 2004. Dialogue and its role in the development of relationship specific knowledge. *Journal of Business & Industrial Marketing*, **19**(2), 114–123.

Barnes, James G. 1997. Closeness, strength, and satisfaction: Examining the nature of relationships between providers of financial services and their retail customers. *Psychology and Marketing*, **14**(8), 765–790.

Batt, Peter. 2004. Managing collaboration within networks and relationships. *Industrial Marketing Management*, **33**(3), 169–174.

Bauer, Kent. 2004. The KPI Profiler: CRM Case Study. *DM Review*.

Beauchamp, Murray A. 1965. An improved index of centrality. *Behavioral Science*, **10**(2), 161–163.

Bennett, Rebekah, & Bove, Liliana L. 2001. Identifying the Key Issues for Measuring Loyalty. *Australasian Journal of Market Research*, **9**, 27–44.

Bennett, Robert, & Robson, Paul. 1999. *Intensity of Interaction in Supply of Business Advice and Client Impact: a Comparison of Consultancy, Business Associations and Government Support Initiatives for SMEs*.

Berger, Paul D., Bolton, Ruth N., Bowman, Douglas, Briggs, Elten, Kumar, V., Parasuraman, A, & Creed, Terry. 2002. Marketing Actions and the Value of Customer Assets: A Framework for Customer Asset Management. *Journal of Service Research*, **5**(1), 39–54.

Berry, Leonard L. 1983. Relationship Marketing. *Pages 25–38 of:* Berry, Leonard T., Shostak, G Lynn, & Upah, Gregory D. (eds), *Emerging Perspectives on Services Marketing*. Chicago: American Marketing Association.

Berry, Leonard L. 1995. Relationship Marketing of Services - Growing Interest, Emerging Perspectives. *Journal of the Academy of Marketing Science*, **23**(4), 236–245.

Berry, Michael J. A., & Linoff, Gordon S. 2004. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. 2 edn. Wiley Computer Publishing.

Bettencourt, Lance A. 1997. Customer voluntary performance: Customers as partners in service delivery. *Journal of Retailing*, **73**(3), 383–406.

Biemans, Wim G., Brencic, Maja, & Malshe, Avinash. 2010. Marketing sales interface configurations in B2B firms. *Industrial Marketing Management*, **39**(2), 183–194.

Blois, Robert. 2008. *CRM Spending up, but User Adoption Still a Problem*. http://www.itbusinessedge.com/cm/community/features/interviews/blog/crm-spending-up-but-user-adoption-still-a-problem/?cs=22697, accessed on 20/11/2011.

Boles, James, Johnston, Wesley, & Gardner, Alston. 1999. The selection and organization of national accounts: a North American perspective. *Journal of Business & Industrial Marketing*, **14**(4), 264–282.

Bolton, Ruth N. 2006. Foreword. *Page ix of:* Lusch, Robert F., & Vargo, Stephen L. (eds), *The Service-Dominant Logic of Marketing*. M.E.Sharpe.

Bove, Liliana L., & Johnson, Lester W. 2001. Customer relationships with service personnel: do we measure closeness, quality or strength? *Journal of Business Research*, **54**(3), 189–197.

Brandes, Ulrik, & Erlebach, Thomas. 2005a. Fundamentals. *Chap. 2, pages 7–15 of:* Brandes, Ulrik, & Erlebach, Thomas (eds), *Network Analysis: Methodological Foundations*. Springer.

Brandes, Ulrik, & Erlebach, Thomas. 2005b. Introduction. *Chap. 1 of:* Brandes, Ulrik, & Erlebach, Thomas (eds), *Network Analysis: Methodological Foundations*. Springer.

Bruhn, Manfred, & Georgi, Dominik. 2006. *Services marketing: managing the service value chain*. 1 edn. Pearson Education.

Buechel, Berno, & Buskens, Vincent. 2008. *The Dynamics of Closeness and Betweenness*.

Bueren, Adrian, Schierholz, Ragnar, Kolbe, Lutz, & Brenner, Walter. 2004. Customer knowledge management - improving performance of customer relationship management with knowledge management. *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, **00**(C), 172–181.

Burton, Betsy, Genovese, Yvonne, Rayner, Nigel, Casonato, Regina, Smith, Michael, Beyer, Mark A, Austin, Tom, Gassman, Bill, & Sommer, Dan. 2011. Predicts 2011 : Pattern-Based Strategy Technologies and Business Practices Gain Momentum. *Gartner Inc.*, 1–15.

Cannon, Joseph P., & Narayandas, Narakesari. 2000. Relationship Marketing and Key Account Management. *Chap. 15, pages 407–429 of:* Sheth, Jagdish N., & Parvatiyar, Atul (eds), *Handbook of Relationship Marketing*, 1 edn. Sage Publications, Inc.

Cermak, Dianne S.P., & File, Karen Maru. 1994. Customer participation in service specification and delivery. *Journal of Applied Business Research*, **10**(2), 90–100.

Chapman, Pete, Clinton, Julian, Kerber, Randy, Khabaza, Thomas, Reinartz, Thomas, Shearer, Colin, & Wirth, Rüdiger. 2000. *Step-by-step data mining guide*. Tech. rept.

Chappell, David. 2010. *Introducing Windows Communication Foundation*. Tech. rept. January. Microsoft Corporation.

Chesbrough, Henry W. 2007. Why Companies Should Have Open Business Models. *MIT Sloan Management Review*, **48**(2), 22–28.

Christopher, Martin, Payne, Adrian, & Ballantyne, David. 1993. *Relationship Marketing: Bringing quality, customer service and marketing together*. Butterworth-Heinemann.

Coase, Ronald H. 1937. The Nature of the Firm. *Economica*, **4**(16), 386–405.

Crosby, Lawrence A., Evans, Kenneth R., & Cowles, Deborah. 1990. Relationship Quality in Services Selling : An Interpersonal Influence Perspective. *Journal of Marketing*, **54**(3), 68–81.

Cuganesan, Suresh. 2008. Calculating customer intimacy: accounting numbers in a sales and marketing department. *Accounting, Auditing & Accountability Journal*, **21**(1), 78–103.

Dalkir, Kimiz. 2011. *Knowledge Management in Theory and Practice*. 2 edn. The MIT Press.

Davenport, Thomas H. 2006. Competing on analytics. *Harvard Business Review*, **84**(1), 98–107, 134.

Davenport, Thomas H., & Harris, Jeanne G. 2007. *Competing on Analytics: The New Science of Winning*. 1 edn. Harvard Business School Press.

Day, George S. 2003. Creating a Superior Customer-Relating Capability. *MIT Sloan Management Review*.

Day, George S., & Montgomery, David B. 1999. Charting New Directions for Marketing. *Journal of Marketing*, **63**(Special Issue 1999), 3–13.

De Choudhury, Munmun, Mason, Winter A., Hofman, Jake M., & Watts, Duncan J. 2010. Inferring relevant social networks from interpersonal communication. *Proceedings of the 19th international conference on World wide web - WWW '10*, 301.

De Jong, Ad, & De Ruyter, Ko. 2004. Adaptive versus Proactive Behavior in Service Recovery: The Role of Self-Managing Teams. *Decision Sciences*, **35**(3), 457–491.

De Wulf, Kristof, Odekerken-Schröder, Gaby, & Iacobucci, Dawn. 2001. Investments in Consumer Relationships : A Cross-Country and Cross-Industry Exploration. *Journal of Marketing*, **65**(October), 33–50.

Den Butter, Frank A. G. 2010. *Transaction Management : Value Creation by Reducing Transaction Costs*.

Deng, Shengliang, & Dart, Hack. 1994. Measuring Market Orientation: A Multi-factor, Multi-item Approach. *Journal of Marketing Management*, **10**(8), 725–742.

Dick, Alan S., & Basu, Kunal. 1994. Customer Loyalty: Toward an Integrated Conceptual Framework. *Journal of the Academy of Marketing Science*, **22**(2), 99–113.

Dickie, Jim. 2007. Fueling the crm engine. *DestinationCRM.com*.

Doherty, Neil F., & Lockett, Nigel J. 2004. Closing the Gap Between the Expectations of Relationship Marketing and the Reality of E-CRM. *International Journal of E-Business Research*, **3**(2), I–VI.

Donaldson, Bill, & O'Toole, Tom. 2007. *Strategic market relationships*. 2nd edn. John Wiley & Sons.

Dunham, Margaret H. 2002. *Data Mining: Introductory and Advanced Topics*. Prentice Hall.

Dyer, Jeffrey H. 1997. Effective interim collaboration: how firms minimize transaction costs and maximise transaction value. *Strategic Management Journal*, **18**(7), 535–556.

Dyer, Jeffrey H., & Chu, Wujin. 2003. The Role of Trustworthiness in Reducing Transaction Costs and Improving Performance: Empirical Evidence from the United States, Japan, and Korea. *Organization Science*, **14**(1), 57–68.

Elzinga, D.Jack, Horak, Tomas, Lee, Chung-Yee, & Bruner, Charles. 1995. Business process management: survey and methodology. *IEEE Transactions on Engineering Management*, **42**(2), 119–128.

Ernerst-Jones, Terry. 2005. *Know how - Managing knowledge for competitive advantage*. Tech. rept. The Economist.

Etzion, Opher, & Niblett, Peter. 2010. *Event Processing in Action.* Manning Publications.

Fayyad, Usama M., Piatetsky-Shapiro, Gregory, & Smyth, Padhraic. 1996a. From Data Mining to Knowledge Discovery: An Overview. *Chap. 1, page 611 of:* Fayyad, Usama M., Piatetsky-Shapiro, Gregory, Smyth, Padhraic, & Uthurusamy, Ramasamy (eds), *Advances in Knowledge Discovery and Data Mining.* AAAI Press / The MIT Presse.

Fayyad, Usama M., Piatetsky-Shapiro, Gregory, & Smyth, Padhraic. 1996b. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, **17**(3), 37–54.

Fornell, Claes, Johnson, Michael D., Anderson, Eugene W., Cha, Jaesung, & Bryant, Barbara Everitt. 1996. The American Customer Satisfaction Index: Nature, Purpose, and Findings. *Journal of Marketing*, **60**(4), 7.

Freeman, L. 1979. Centrality in social networks conceptual clarification. *Social Networks*, **1**(3), 215–239.

Frese, Michael, Fay, Doris, Hilburger, Tanja, Leng, Karena, & Tag, Almut. 1997. The concept of personal initiative: Operationalization, reliability and validity in two German samples. *Journal of occupational and organizational psychology*, **70**(2), 139–161.

Gartner. 2010. *Hype Cycle for Analytic Applications.* Tech. rept. October.

Gebert, Henning, Geib, Malte, Kolbe, Lutz, & Brenner, Walter. 2003. Knowledge-enabled customer relationship management: integrating customer relationship management and knowledge management concepts. *Journal of Knowledge Management*, **7**(5), 107–123.

George, Michael L. 2003. *Lean Six Sigma for Service : How to Use Lean Speed and Six Sigma Quality to Improve Services and Transactions.* McGraw-Hill.

Gibbert, Michael, Leibold, Marius, & Probst, Gilbert. 2002. Five styles of Customer Knowledge Management, and how smart companies put them into action. *European Management Journal*, **20**, 459–469.

Granovetter, Mark. 1985. Economic Action and Social Structure : The Problem of Embeddedness. *The American Journal of Sociology*, **91**(3), 481–510.

Grönroos, Christian. 1994. From Marketing Mix to Relationship Marketing: Towards a Paradigm Shift in Marketing. *Management Decision*, **32**(2), 4–20.

Grönroos, Christian. 1997. Value-driven relational marketing: From products to resources and competencies. *Journal of Marketing Management*, **13**(5), 407–419.

Grönroos, Christian. 2000. Relationship Marketing - The Nordic School Perspective. *Chap. 4, pages 95–117 of:* Sheth, Jagdish N., & Parvatiyar, Atul (eds), *Handbook of Relationship Marketing*, 1 edn. Sage Publications, Inc.

Grönroos, Christian. 2007. *Service Management and Marketing: Customer Management in Service Competition*. 3 edn. Wiley.

Gruen, Thomas W. 1997. Relationship marketing: the route to marketing efficiency and effectiveness. *Business Horizons*, **40**(6), 32–38.

Gummesson, Evert. 1995. *Relationsmarknadsföring: Från 4P till 30R*.

Gummesson, Evert. 1996. Relationship marketing and imaginary organizations: a synthesis. *European Journal of Marketing*, **30**(2), 31–44.

Gummesson, Evert. 2008. *Total Relationship Marketing*. Butterworth-Heinemann.

Gwinner, Kevin P., Bitner, Mary Jo, Brown, Stephen W., & Ajith, Kumar. 2005. Service Customization Through Employee Adaptiveness. *Journal of Service Research*, **8**(2), 131–148.

Habryn, François, Blau, Benjamin, Satzger, Gerhard, & Kölmel, Bernhard. 2010. Towards a Model for Measuring Customer Intimacy in B2B Services. *Pages 1–14 of:* Morin, Jean-Henry, Ralyté, Jolita, & Snene, Mehdi (eds), *Exploring Services Science*. Springer Berlin Heidelberg.

Hair, Joseph F., Black, William C., Babin, Barry J., & Anderson, Rolph E. 2010. *Multivariate Data Analysis (7th Edition)*.

Hakansson, Hakan, & Snehota, Ivan. 2000. The IMP Perspective: Assets and Liabilities of Business Relationships. *Chap. 3, pages 69–93 of:* Sheth, Jagdish N., & Parvatiyar, Atul (eds), *Handbook of Relationship Marketing*. Sage Publications, Inc.

Hakansson, Hakan, Ford, David I., Gadde, Lars-Erik, Snehota, Ivan, & Waluszewski, Alexandra. 2009. *Business in Networks*. Wiley.

Hambrick, Donald C. 1980. Operationalizing the Concept of Business-Level Strategy in Research. *Academy of Management Journal*, **5**(4), 567–575.

Harding, Ford. 2004. Cross Selling or Cross Purposes ? *Harvard Business Review Case Study*.

Hasan, Matt. 2003. Ensure success of CRM with a change in mindset. *Marketing News*, **37**(8), 16.

Hennig-Thurau, Thorsten, Gwinner, Kevin P., & Gremler, Dwayne D. 2002. Understanding Relationship Marketing Outcomes: An Integration of Relational Benefits and Relationship Quality. *Journal of Service Research*, **4**(3), 230–247.

Heskett, James L, Jones, Thomas O, Loveman, Gary W, Sasser, W Earl, & Schlesinger, Leonard A. 1994. Putting the Service-Profit Chain to Work. *Harvard Business Review*, **72**(2), 164–170.

Hodnick, Mike. 2009. *Connectable Lines and Shapes in Silverlight*.

Holmlund, Maria. 2004. Analyzing business relationships and distinguishing different interaction levels. *Industrial Marketing Management*, **33**(4), 279–287.

Holt, Sue, & McDonald, Malcolm. 2000. Managing Global Networks : The Role of the Global Account Manager. *Pages 1–16 of: 16th IMP Conference*.

Homburg, Christian, & Jensen, Ove. 2004. *Kundenbindung im Industriegütergeschäft*. Universität Mannheim - Institut für Marktorientierte Unternehmensführung; Auflage: 1., Aufl.

Husted, Bryan W. 1994. Transaction Costs, Norms, and Social Networks: A Preliminary Study of Cooperation in Industrial Buyer-Seller Relationships in the United States and Mexico. *Business & Society*, **33**(1), 30–57.

Hutt, Michael D., & Walker, Beth A. 2006. A network perspective of account manager performance. *Journal of Business & Industrial Marketing*, **21**(7), 466–473.

IBM Institute for Business Value. 2010. *Capitalizing on Complexity - Insight from the Global CEO Study*. Tech. rept.

IBM Institute for Business Value. 2011. *Insights from the Global Chief Marketing Officer Survey - From Stretched to Strenghened*. Tech. rept.

Industry Directions, & IBM. 2006. *Increasing Customer Intimacy, Value-Add and Profits Through Expanded Services*. Tech. rept.

Jamieson, Susan. 2004. Likert scales: how to (ab)use them. *Medical education*, **38**(12), 1217–8.

Jaworski, Bernard J., & Kohli, Ajay K. 1993. Market Orientation: Antecedents and Consequences. *The Journal of Marketing*, **57**(3), 53–70.

Jayachandran, Satish, Hewett, Kelly, & Kaufman, Peter. 2004. Customer Response Capability in a Sense-and-Respond Era: The Role

of Customer Knowledge Process. *Journal of the Academy of Marketing Science*, **32**(3), 219–233.

John, George H. 1995. Robust Decision Trees : Removing Outliers from Databases. *Pages 174–179 of:* Fayyad, Usama M., & Uthurusamy, R. (eds), *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*. Menlo Park: AAAI Press.

Joshi, Ashwin W., & Sharma, Sanjay. 2004. Customer Knowledge Development : Antecedents and Impact on New Product Performance. *Journal of Marketing*, **68**(October), 47–59.

Kale, Sudhir H. 2004. CRM Failure and the Seven Deadly Sins. *Marketing Management*, **13**(5), 42–46.

Kalousis, Alexandros, Gama, João, & Hilario, Melanie. 2004. On Data and Algorithms: Understanding Inductive Performance. *Machine Learning*, **54**(3), 275–312.

Kamakura, Wagner A., Ramaswamy, Sridhar N., & Srivastava, Rajendra K. 1991. Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. *International Journal of Research in Marketing*, **8**(4), 329–349.

Kamakura, Wagner A., Wedel, Michel, De Rosa, Fernando, & Mazzon, Jose Afonso. 2003. Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction. *International Journal of Research in Marketing*, **20**(1), 45–65.

Kaplan, Robert S. 2005. *A Balanced Scorecard Approach To Measure Customer Profitability*.

Kaplan, Robert S., & Anderson, Steven R. 2007. *Time-Driven Activity-Based Costing: A Simpler and More Powerful Path to Higher Profits*. Harvard Business School Press.

Kirkman, Bradley L., Rosen, Benson, Tesluk, Paul E., & Gibson, Cristina B. 2004. The Impact of Team Empowerment on Virtual Team Performance: The Moderating Role of Face-to-Face Interaction. *The Academy of Management Journal*, **47**(2), 175–192.

Kiron, David, Shockley, Rebecca, Kruschwitz, Nina, Finch, Glenn, & Haydock, Michael. 2011. *Analytics : The widening divide*. Tech. rept. IBM Institute for Business Value.

Kiss, Christine. 2007. *Leveraging social networks for analytical customer relationship management*. Ph.D. thesis.

Knoke, David, & Yang, Sang. 2008. *Social Network Analysis*. 2 edn. Sage Publications, Inc.

Kohavi, Ron. 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Pages 1137–1143 of: Proceedings of the 14th International Joint Conference on Artificial Intelligence*.

Kohli, Ajay K., & Jaworski, Bernard J. 1990. Market Orientation : The Construct, Research Propositions Managerial Implications. *Journal of Marketing*, **54**(2), 1–18.

Kong, Raymond, & Mayo, Michael C. 1993. Measuring Service Quality in the Business-to-business Context. *Journal of Business and Industrial Marketing*, **8**(2), 5–15.

Koschützki, Dirk, Lehmann, Katharina Anna, Peeters, Leon, Richter, Stefan, Tenfelde-Podehl, Dagmar, & Zlotowski, Oliver. 2005. Centrality Indices. *Chap. 3, pages 16–61 of:* Brandes, Ulrik, & Erlebach, Thomas (eds), *Network Analysis: Methodological Foundations*. Springer.

Kumar, Ashok. 2005. Mass Customization: Metrics and Modularity. *International Journal of Flexible Manufacturing Systems*, **16**(4), 287–311.

Lages, Carmen, Lages, Cristiana Raque, & Lages, Luis Filipe. 2005. The RELQUAL scale: a measure of relationship quality in export market ventures. *Journal of Business Research*, **58**(8), 1040–1048.

Lam, Shun Y., Shankar, Venkatesh, Erramilli, M. Krishna, & Murthy, Bvsan. 2004. Customer Value, Satisfaction, Loyalty, and Switching

Costs: An Illustration From a Business-to-Business Service Context. *Journal of the Academy of Marketing Science*, **32**(3), 293–311.

Lam, Wai, Keung, Chi-kin, & Liu, Danyu. 2002. Discovering useful concept prototypes for classification based on filtering and abstraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**(8), 1075–1090.

Leek, Sheena, Turnbull, Peter W., & Naudé, Pete. 2001. Interactions , Relationships and Networks - Past , Present and Future. *Pages 1–13 of: 17th IMP-conference*, vol. 1.

Liang, Ting-peng. 2009. Information Technology for Customer Intimacy : A Niche for Research in the Internet Age. *Pacific Asia Journal of the Association for Information Systems*, **1**(3), 1–4.

Liljander, Veronika, & Strandvik, Tore. 1995. The Nature of Relationships in Services. *Advances in Service Marketing and Management*, **4**, 143.

Lovelock, Christopher H., & Wirtz, Jochen. 2007. *Services Marketing*. 6., inter. edn. Pearson Prentice Hall.

Magnusson, Peter R., Matthing, Jonas, & Kristensson, Per. 2003. Managing User Involvement in Service Innovation: Experiments with Innovating End Users. *Journal of Service Research*, **6**(2), 111–124.

Malms, Oliver, & Schmitz, Christian. 2011. Cross-Divisional Orientation: Antecedents and Effects on Cross-Selling Success. *Journal of Business-to-Business Marketing*, **18**(3), 253–275.

Manasco, Britton. 2000. Capitalizing on Customer Knowledge. *APQC*, 66–70.

Maoz, Michael, Davies, Jim, Thompson, Ed, Dunne, Michael, & Sarner, Adam. 2010. *The Concise Social CRM Vendor Guide , 2011*. Tech. rept. October 2010. Gartner, Inc.

McDonald, Malcolm, Millman, Tony, & Rogers, Beth. 1997. Key account management: Theory, practice and challenges. *Journal of Marketing Management*, **13**(8), 737–757.

Medlin, Christopher John, & Törnroos, Jan-Ake. 2006. Inter-firm Interaction from a Human Perspective. *In: 22nd IMP-conference.*

Michie, Donald, Spiegelhalter, David J., & Taylor, Charles C. 1994. Machine Learning, Neural, and Statistical Classification. *Journal of the American Statistical Association*, **91**(433), 436– 438.

Miller, Delbert C., & Salkind, Neil J. 2002. *Handbook of Research Design and Social Measurement*. 6th edn. Sage Publications, Inc.

Mitchell, James C. 1969. *Social Network in Urban Situations: analyses of personal relationships in Central African towns.*

Mitchell, Tom M. 1999. Machine Learning and Data Mining Over the past. *Communications of the ACM*, **42**(11).

Mohr, Jakki, & Nevin, John R. 1990. Communication Strategies in Marketing Channels : A Theoretical Perspective. *The Journal of Marketing*, **54**(4), 36–51.

Morgan, Robert M., & Hunt, Shelby D. 1994. The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing*, **58**(3), 20– 38.

Narver, John C., & Slater, Stanley F. 1990. The Effect of a Market Orientation on a Market Business Profitability. *The Journal of Marketing*, **54**(4), 20–35.

Nash, Debbie, & Sterna-Karwat, Alicia. 1996. An application of DEA to measure branch cross selling efficiency. *Computers & Operations Research*, **23**(4), 385–392.

Ndubisi, Nelson Oly. 2007. Relationship marketing and customer loyalty. *Marketing Intelligence & Planning*, **25**(1), 98–106.

Neely, Andy. 2009. Exploring the financial consequences of the servitization of manufacturing. *Operations Management Research*, **1**(2), 103–118.

Newman, Mark E.J. 2004. Analysis of weighted networks. *Physical Review E*, **70**(5), 1–9.

Nezlek, John B. 2003. Using Multilevel Random Coefficient Modeling to Analyze Social Interaction Diary Data. *Journal of Social and Personal Relationships*, **20**(4), 437–469.

Nisbet, Robert, Elder, John, & Miner, Gary. 2009. *Handbook of Statistical Analysis and Data Mining Applications*. Academic Press.

Niven, Paul R. 2002. *Balanced Scorecard Step-by-Step: Maximizing Performance and Maintaining Results*. Wiley.

Noorderhaven, Niels, & Harzing, Anne-Wil. 2009. Knowledge-sharing and social interaction within MNEs. *Journal of International Business Studies*, **40**(5), 719–741.

Odekerken-Schröder, Gaby, De Wulf, Kristof, & Schumacher, Patrick. 2003. Strengthening outcomes of retailer-consumer relationships. The dual impact of relationship marketing tactics and consumer personality. *Journal of Business Research*, **56**(3), 177–190.

Oliva, Rogelio, & Kallenberg, Robert. 2003. Managing the transition from products to services. *International Journal of Service Industry Management*, **14**(2), 160–172.

Oliver, Richard L. 1999. Whence Consumer Loyalty? *Journal of Marketing*, **63**(1999), 33.

Ostaszewski, Adam. 1990. *Advanced mathematical methods*. Cambridge University Press.

Palmatier, Robert W., Dant, Rajiv P., Grewal, Dhruv, & Evans, Kenneth R. 2006. Factors Influencing the Effectiveness of Relationship Marketing: A Meta-Analysis. *Journal of Marketing*, **70**(October), 136–153.

Palmatier, Robert W., Scheer, Lisa K., & Steenkamp, Jan-Benedict E.M. 2007. Customer Loyalty to Whom? Managing the Benefits and Risks of Salesperson-Owned Loyalty. *Journal of Marketing Research*, **44**(2), 185–199.

Parvatiyar, Atul, & Sheth, Jagdish N. 2000. The Domain and Conceptual Foundations of Relationship Marketing. *Chap. 1, pages 3–38 of:* Sheth, Jagdish N., & Parvatiyar, Atul (eds), *Handbook of Relationship Marketing*, 1 edn. Sage Publications, Inc.

Parvatiyar, Atul, & Sheth, Jagdish N. 2001. Conceptual Framework of Customer Relationship Management. *Chap. 1, pages 3–25 of:* Sheth, Jagdish N., Parvatiyar, Atul, & Shainesh, G (eds), *Customer relationship management: emerging concepts, tools, and applications*, 1 edn. Tata McGraw-Hill.

Patton, Michael Q. 2002. *Qualitative research and evaluation methods*. 3 edn. Sage Publications, Inc.

Payne, Adrian. 2005. *Handbook of CRM: Achieving Excellence through Customer Management*. Butterworth-Heinemann.

Payne, Adrian, & Frow, Pennie. 2005. A Strategic Framework for Customer Relationship Management. *Journal of Marketing*, **69**(4), 167–176.

Plinke, Wulff. 1997. Grundlagen des Geschäftsbeziehungsmanagement. *Chap. 1, pages 1–62 of:* Kleinaltenkamp, Michael, & Plinke, Wulff (eds), *Geschäftsbeziehungsmanagement*. Springer.

Porter, Michael E. 2004. *Competitive Strategy*. Free Press.

Potgieter, A, & Roodt, G. 2004. Measuring a customer intimacy culture in a value discipline context. *South African Journal of Human Resource Management*, **2**(3), 25–31.

Press, S. James. 2003. The Role of Bayesian and Frequentist Multivariate Modeling in Statistical Data Mining. *Chap. 1, page 624 of:* Bozdogan, Hamparsum (ed), *Statistical Data Mining and Knowledge Discovery*. Chapman and Hall/CRC.

Quinlan, John R. 1986. Induction of decision trees. *Machine Learning*, **1**(1), 81–106.

Reichheld, Frederick F. 2003. The one number you need to grow. *Harvard Business Review*, **81**(12), 46–54, 124.

Reichheld, Frederick F., & Sasser, W. Earl. 1990. Zero defections: quality comes to services. *Harvard Business Review*, **68**(5), 105–111.

Reichheld, Frederick F., & Teal, Thomas. 2001. *The loyalty effect: the hidden force behind growth, profits, and lasting value*. Harvard Business Press.

Reinartz, Werner J., & Kumar, V. 2000. On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing. *Journal of Marketing*, **64**(4), 17–35.

Reinartz, Werner J., Krafft, Manfred, & Hoyer, Wayne D. 2004. The Customer Relationship Management Process: Its Measurement and Impact on Performance. *Journal of Marketing Research*, **41**(3), 293–305.

Richard, James E. 2008. *The Impact of Customer Relationship Management (CRM) Technology on Business-to-Business Customer Relationships*. Ph.D. thesis, Victoria University of Wellington.

Richards, Keith A., & Jones, Eli. 2008. Customer relationship management: Finding value drivers. *Industrial Marketing Management*, **37**(2), 120–130.

Richardson, Leonard, & Ruby, Sam. 2007. *Restful Web Services*. 1st edn. O'Reilly Media, Inc, USA.

Roberts, Keith, Varki, Sajeev, & Brodie, Rod. 2003. Measuring the quality of relationships in consumer services: an empirical study. *European Journal of Marketing*, **37**(1/2), 169–196.

Robinson, John P., Shaver, Phillip R., & Wrightsman, Lawrence S. 1991. Criteria for Scale Selection and Evaluation. *Chap. 1, pages*

1–16 of: Robinson, John P, Shaver, Phillip R, & Wrightsman, Lawrence S (eds), *Measures of personality and social psychological attitudes*. Elsevier.

Rodrigez Cano, Cynthia, Carillat, Francois A., & Jaramillo, Fernando. 2004. A meta-analysis of the relationship between market orientation and business performance: evidence from five continents. *International Journal of Research in Marketing*, **21**(2), 179–200.

Rust, Roland T., Lemon, Katherine N., & Zeithaml, Valarie A. 2004. Return on Marketing: Using Customer Equity to Focus Marketing Strategy. *Journal of Marketing*, **68**(1), 109–127.

Ryals, Lynette, & Knox, Simon. 2001. Cross-functional issues in the implementation of relationship marketing through customer relationship management. *European Management Journal*, **19**(5), 534–542.

Sabidussi, Gert. 1966. The centrality index of a graph. *Psychometrika*, **31**(4), 581–603.

Sackmann, Stefan, Kundisch, Dennis, & Ruch, Markus. 2008. CRM, Kundenbewertung und integrierte Risk-Return-Steuerung - Zum Status quo des betrieblichen Einsatzes. *HMD Praxis der Wirtschaftsinformatik*, **44**(259), 21–32.

Sako, Mari. 1992. *Prices, quality, and trust: inter-firm relations in Britain and Japan*.

Sandberg, Birgitta. 2007. Customer-related proactiveness in the radical innovation development process. *European Journal of Innovation Management*, **10**(2), 252–267.

Satzger, Gerhard, & Neus, Andreas. 2010. Principles of Collaborative Innovation: Implementation of Open and Peer-to-Peer Innovation Approaches. *Pages 219–238 of:* Gerybadze, Alexander, Hommel, Ulrich, Reiners, Hans W., & Thomaschewski, Dieter (eds), *Innovation and International Corporate Growth*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Shapiro, Benson P., & Wyman, John. 1981. New Ways to Reach Your Customers. *Harvard Business Review*, **59**(6), 103–110.

Sharma, Neeru, & Patterson, Paul G. 1999. The impact of communication effectiveness and service quality on relationship commitment in consumer, professional services. *Journal of Services Marketing*, **13**(2), 151–170.

Sheth, Jagdish N., & Parvatiyar, Atul. 2000. The evolution of relationship marketing. *Chap. 5, pages 119–145 of:* Sheth, Jagdish N., & Parvatiyar, Atul (eds), *Handbook of Relationship Marketing*, vol. 4. Sage Publications, Inc.

Sheth, Jagdish N., & Sharma, Arun. 2008. The impact of the product to service shift in industrial markets and the evolution of the sales organization. *Industrial Marketing Management*, **37**(3), 260–269.

Silveira, Da Giovanni, Borenstein, Denis, & Fogliatto, Flávio S. 2001. Mass customization: Literature review and research directions. *International Journal of Production Economics*, **72**(1), 1–13.

Sommerville, Ian. 2007. *Software Engineering*. 8th edn. Pearson Education.

Stern, Barbara B. 1997. Advertising Intimacy : Relationship Marketing and the Services Consumer. *Journal of Advertizing*, **26**(4), 7–19.

Storbacka, Kaj, Strandvik, Tore, & Grönroos, Christian. 1994. Managing Customer Relationships for Profit: The Dynamics of Relationship Quality. *International Journal of Service Industry Management*, **5**(5), 21–38.

Tan, Pang-Ning, Steinbach, Michael, & Kumar, Vipin. 2006. *Introduction to Data Mining*. Addison Wesley.

Treacy, Michael, & Wiersema, Frederik D. 1993. Customer intimacy and other value disciplines. *Harvard Business Review*.

Treacy, Michael, & Wiersema, Frederik D. 1997. *The discipline of market leaders*. Basic Books.

Tsvetovat, Maksim, Reminga, Jeff, & Carley, Kathleen M. 2004. *DyNetML : Interchange Format for Rich Social Network Data*. Tech. rept. February. Carnegie Mellon University - School of Computer Science, Pittsburg.

Tu, Y. L., Xie, S. Q., & Fung, Richard Y. K. 2007. Product Development Cost Estimation in Mass Customization. *IEEE Transactions on Engineering Management*, **54**(1), 29–40.

Tuominen, Matti, Rajala, Arto, & Mo, Kristian. 2004. Market-driving versus market-driven: Divergent roles of market orientation in business relationships. *Industrial Marketing Management*, **33**(3), 207–217.

Turban, Efraim, Sharda, Ramesh, & Delen, Dursun. 2011. *Decision Support and Business Intelligence Systems (International Edition)*. 9th edn. Pearson Education.

Vandermerwe, Sandra. 1988. Servitization of business: Adding value by adding services. *European Management Journal*, **6**(4), 314–324.

Varadarajan, P. Rajan, & Cunningham, Margaret H. 1995. Strategic Alliances: A Synthesis of Conceptual Foundations. *Journal of the Academy of Marketing Science*, **23**(4), 282–296.

Varadarajan, P. Rajan, & Rajaratnam, Daniel. 1986. Symbiotic Marketing Revisited. *Journal of Marketing*, **50**(1), 7.

Vargo, Stephen L., & Lusch, Robert F. 2004a. Evolving to a New Dominant Logic for Marketing. *Journal of Marketing*, **68**(1), 1–17.

Vargo, Stephen L., & Lusch, Robert F. 2004b. The Four Service Marketing Myths: Remnants of a Goods-Based, Manufacturing Model. *Journal of Service Research*, **6**(4), 324–335.

Vargo, Stephen L., & Lusch, Robert F. 2006. Service-Dominant Logic - What It Is, What It Is Not, What It Might Be. *Chap. 3, pages 43–56 of:* Vargo, Stephen L., & Lusch, Robert F. (eds), *The Service-Dominant Logic of Marketing*, 1 edn. M.E.Sharpe.

Vargo, Stephen L., & Lusch, Robert F. 2008a. Service-dominant logic: continuing the evolution. *Journal of the Academy of Marketing Science*, **36**(1), 1–10.

Vargo, Stephen L., & Lusch, Robert F. 2008b. Why "service"? *Journal of the Academy of Marketing Science*, **36**(1), 25–38.

Venkatesan, Rajkumar, & Kumar, V. 2004. A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. *Journal of Marketing*, **68**(4), 106–125.

Wallenburg, Carl Marcus. 2009. Innovation in Logistics Outsourcing Relationships: Proactive Improvement by Logistics Service Providers as a Driver of Customer Loyalty. *Journal of Supply Chain Management*, **45**(2), 75–93.

Wallenburg, Carl Marcus, Knemeyer, A. Michael, Goldsby, Thomas J., & Cahill, David L. 2010. Developing a scale for proactive improvement within logistics outsourcing relationships. *The International Journal of Logistics Management*, **21**(1), 5–21.

Wasserman, Stanley, & Faust, Katherine. 1994. *Social Network Analysis: Methods and Applications*. 1 edn. Cambridge University Press.

Weill, Peter, & Ross, Jeanne W. 2004. *IT governance: how top performers manage IT decision rights for superior results*. Harvard Business School Press.

Wengler, Stefan. 2006. *Key account management in business-to-business markets*. Ph.D. thesis.

Wengler, Stefan, Ehret, Michael, & Saab, Samy. 2006. Implementation of Key Account Management: Who, why, and how?: An exploratory study on the current implementation of Key Account Management programs. *Industrial Marketing Management*, **35**(1), 112–103.

Williamson, Oliver E. 1979. Transaction-Cost Economics: The Governance of Contractual Relations. *Journal of Law and Economics*, **22**(2), 233–261.

Wind, Jerry, & Rangaswamy, Arvind. 2001. Customerization: The next revolution in mass customization. *Journal of Interactive Marketing*, **15**(1), 13–32.

Witten, Ian H., Eibe, Frank, & Hall, Mark A. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. 3 edn. Morgan Kaufmann.

Wu, Lynn, Lin, Ching-yung, Stern, N Y U, & Brynjolfsson, Erik. 2009. Value of Social Network – A Large-Scale Analysis on Network Structure Impact to Financial Revenue of Information Technology Consultants. *In: Winter Information Systems Conference*.

Wuyts, Stefan, & Geyskens, Inge. 2005. The Formation of Buyer-Supplier Relationships: Detailed Contract Drafting and Close Partner Selection. *Journal of Marketing*, **69**(4), 103–117.

Yim, Chi Kin, Tse, David K, & Chan, Kimmy Wa. 2008. Strengthening Customer Loyalty Through Intimacy and Passion: Roles of Customer-Firm Affection and Customer-Staff Relationships in Services. *Journal of Marketing Research*, **45**(6), 741–756.

Zablah, Alex, Bellenger, Danny N., & Johnston, Wesley J. 2004. An evaluation of divergent perspectives on customer relationship management: Towards a common understanding of an emerging phenomenon. *Industrial Marketing Management*, **33**(6), 475–489.

Zack, Michael, McKeen, James, & Singh, Satyendra. 2009. Knowledge management and organizational performance: an exploratory analysis. *Journal of Knowledge Management*, **13**(6), 392–409.

Zajac, Edward J., & Olsen, Cyrus P. 1993. From Transaction Cost to Transactional Value Analysis: Implications for the Study of Interorganizational Strategies. *Journal of Management Studies*, **30**(1), 131–145.

Appendix

A. Questionnaire Customer Intimacy

This appendix presents the questionnaire conceived in the scope of this thesis in order to perform the empirical assessment of the customer intimacy components. This questionnaire consists of four different parts:

1. **Introduction:** This section introduces the scope of the survey

2. **Acquired Customer Intimacy – Organization Level:** In this part of the questionnaire, the acquired customer intimacy components at the organizational level are empirically assessed on Likert-type scales with a set of four items.

3. **Acquired Customer Intimacy – Individual Level:** In this part of the questionnaire, the acquired customer intimacy components at the individual level are empirically assessed with a set of four items.

4. **Work Environment:** Finally, in this part of the questionnaire, the respondents are asked to provide further information on their work environment. This part consists of 11 items.

A.1. English Version

In this section, the English version of the questionnaire is presented

SURVEY

Research Project on Customer Relationships

in partnership with

CAS Software AG

Karlsruhe Service Research Institute (KSRI) - Karlsruhe Institute of Technology (KIT)

Questionnaire prepared for: Max Mustermann

Date: 15. September 2010

Thank you for participating in this survey !

Your participation is on a voluntary basis

This will take approximately 20 minutes of your time

Please return your answers by Friday October 5th to Francois Habryn at francois.habryn@kit.edu

Why this research ?

We are interested to know how knowledge of and relationship with customers benefit the competitive advantage of CAS.

This project is performed under the supervision of CAS Directors Dr. Bernhard Kölmel and Spiros Alexakis.

We will ensure that your answers are handled on a strictly confidential basis and only for the purpose of this research project. KSRI will anonymize the data prior to any analysis.

If you are interested, we will be happy to share the results of our research with you.

What is the content of this questionnaire ?

This questionnaire is tailored for you. It is composed of four parts:

1. Your relationship with and knowledge of customer organizations you have worked with

2. Your relationship with and knowledge of employees of these customer organizations

3. Other characteristics of your relationship with these organizations

4. Information on your work environment

What do you have to do?

1. Save this file on your computer

2. Answer the questions by selecting the corresponding option in the dropdown menu

3. Send the file with your answers via e-Mail to francois.habryn@kit.edu

Thank you for your participation to this project !

Figure A.1.: Customer Intimacy Questionnaire: Introduction

Part 1. Please rate your knowledge of and your relationship with the following customers.

Click on "Select Your Answer", then type a number between 1 (strongly disagree) and 7 (strongly agree) or select a value in the dropdown menu.*

		strongly disagree			neutral			strongly agree
		1	2	3	4	5	6	7

	Customer Organization	My knowledge of this customer's needs is thorough	I learned a lot about this customer's preferences in the period I worked with it.	I know this customer very well	As an employee, I have a high-quality relationship with this customer	As an employee, I have a very collaborative relationship with this customer	I am satisfied with the relationship I have with this customer
1	Customer 1	7	3	3	4	3	3
2	Customer 2	6	7	7	7	7	7

Figure A.2.: Customer Intimacy Questionnaire: Acquired Customer Intimacy at the Organizational Level

Part 2. Please rate your knowledge of and your relationship with the following customer employees.

Click on "Select Your Answer", then type a number between 1 (strongly disagree) and 7 (strongly agree) or select a value in the dropdown menu.*

strongly disagree			neutral			strongly agree
1	2	3	4	5	6	7

	Customer Organization	Customer Employee Surname	Customer Employee Firstname	My knowledge of this employee's needs is thorough	I learned a lot about this employee's preferences in the period I worked with him/her	I have a high-quality relationship with this employee.	I have a very collaborative relationship with this employee
1	Customer 1	A	Matthias	Select Your Answer	Select Your Answer	Select Your Answer	Select Your Answer
2	Customer 2	B	Silke	4	3	4	3
3	Customer 3	C	Andreas	2	2	2	2
4	Customer 1	D	Andreas	2	3	4	3
5	Customer 2	E	Wassili	2	4	4	3
6	Customer 2	F	Markus	6	7	6	7
7	Customer 2	G	Herta	6	3	5	5

Figure A.3.: Customer Intimacy Questionnaire: Acquired Customer Intimacy at the Individual Level

Figure A.4.: Customer Intimacy Questionnaire: Work Environment

A.2. German Version

Since the respondents of the survey are from Germany, the customer intimacy questionnaire has been translated in the German language. This section presents this translated questionnaire.

UMFRAGE

Forschungsprojekt im Bereich Customer Relationship

in Kooperation mit

CAS Software AG

Karlsruhe Service Research Institute (KSRI) - Karlsruhe Institute of Technology (KIT)

Fragebogen für: Max Mustermann
Datum: 15. September 2010

Vielen Dank für Ihre Teilnahme an der Umfrage!

Ihre Teilnahme ist freiwillig.

Die Umfrage wird ca. 20 Minuten dauern.

Bitte senden Sie Ihren ausgefüllten Fragebogen bis spätestens Donnerstag, den 9. Dezember, an François Habryn, francois.habryn@kit.edu.

Über das Forschungsprojekt

Uns interessiert wie Wissen über und Beziehungen mit Kunden der CAS helfen können, um einen Wettbewerbsvorteile zu erreichen.

Das Projekt wird unter der Aufsicht von den CAS Direktoren Dr. Bernhard Kölmel und Spiros Alexakis durchgeführt.

Wir versichern Ihnen, dass Ihre Antworten streng vertraulich und nur zum Zwecke dieses Forschungsprojektes verwendet werden. Das KSRI wird die Daten anonymisieren bevor diese analysiert werden.

Falls Sie interessiert sind, informieren wir Sie gerne über die Ergebnisse unseres Forschungsprojektes.

Welchen Inhalt hat der Fragebogen?

Dieser Fragebogen wurde speziell für Sie generiert. Er besteht aus mindestens zwei bis maximal vier Teilen:

1. Ihr Wissen über die und Ihre Beziehungen mit Unternehmen mit denen Sie zusammengearbeitet haben

2. Ihr Wissen über die und Ihre Beziehungen mit den Angestellten dieser Unternehmen

3. Weitere Charakteristika Ihrer Beziehungen mit diesen Unternehmen

4. Informationen über Ihr Arbeitsumfeld

Was müssen Sie machen?

1. Speichern Sie diese Datei auf Ihrem Computer

2. Beantworten Sie die Fragen, indem Sie die entsprechenden Antworten auswählen

3. Senden Sie Ihre beantwortete Excel-Datei per E-Mail an francois.habryn@kit.edu

Vielen Dank für Ihre Teilnahme an diesem Projekt!

Figure A.5.: Customer Intimacy Questionnaire: Introduction (German)
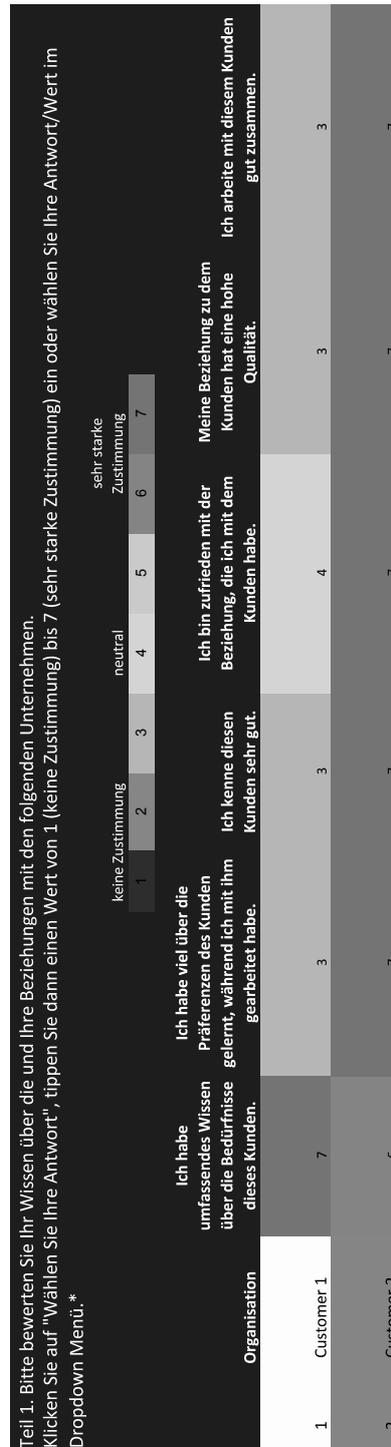
Figure A.6.: Customer Intimacy Questionnaire: Acquired Customer Intimacy at the Organizational Level (German)

Teil 2. Bitte bewerten Sie Ihr Wissen über die und Ihre Beziehungen mit den folgenden Angestellten.
Klicken Sie auf "Wählen Sie Ihre Antwort", tippen Sie dann einen Wert von 1 (keine Zustimmung) bis 7 (sehr starke Zustimmung) ein oder wählen Sie Ihre Antwort/Wert im Dropdown Menü.*

keine Zustimmung			neutral		sehr starke Zustimmung	
1	2	3	4	5	6	7

	Organisation	Nachname des Angestellten	Vorname des Angestellten	Ich kenne diesen Angestellten sehr gut.	Ich habe viel über die Präferenzen des Angestellten gelernt, während ich mit ihr/ihm gearbeitet habe.	Meine Beziehung zu dem Angestellten hat eine hohe Qualität.	Ich arbeite mit diesem Angestellten gut zusammen.
1	Customer 1	A	Matthias	Wählen Sie Ihre Antwort	Wählen Sie Ihre Antwort	Wählen Sie Ihre Antwort	Wählen Sie Ihre Antwort
2	Customer 2	B	Silke	4	3	4	3
3	Customer 3	C	Andreas	2	2	2	2
4	Customer 1	D	Andreas	2	3	4	3
5	Customer 2	E	Wassili	2	4	4	3
6	Customer 2	F	Markus	6	7	6	7
7	Customer 2	G	Herta	6	3	5	5

Figure A.7.: Customer Intimacy Questionnaire: Acquired Customer Intimacy at the Individual Level (German)

Teil 4. Bitte beantworten Sie die folgenden Fragen zu Ihrer Arbeitsumgebung.
Klicken Sie auf "Wählen Sie Ihre Antwort", tippen Sie dann einen Wert von 1 (keine Zustimmung) bis 7 (sehr starke Zustimmung) ein oder wählen Sie Ihre Antwort/Wert im Dropdown Menü. *

		keine Zustimmung			neutral		sehr starke Zustimmung	
		1	2	3	4	5	6	7
1	Ich habe die Freiheit selbst zu entscheiden, was ich in meiner Arbeit mache							7
2	Es liegt im Grunde in meiner Verantwortung, wie ich meine Arbeit mache							7
3	Ich kann frei entscheiden, wie meine Arbeit gemacht wird							7
4	Die Erfolge von CAS sind meine Erfolge							7
5	Wenn jemand CAS lobt, empfinde ich das als persönliches Kompliment							7
6	Wenn CAS in den Medien kritisiert werden würde, wäre ich verlegen							7

Um die folgenden Fragen zu beantworten, können Sie ausschließlich das Dropdown Menü verwenden.

7	Wie alt sind Sie?	2. 31-40
8	Welche Rolle haben Sie bei CAS?	2. Service
9	Welche Position haben Sie bei CAS?	2. Manager
10	Wie lange arbeiten Sie schon bei CAS?	3. zwischen 5 und 10 years
11	Möchten Sie Informationen über die Ergebnisse dieses Forschungsprojektes erhalten?	Ja

Sie haben das Ende der Umfrage erreicht. Vielen Dank für Ihre wertvollen Antworten.

*Anzeigen des Dropdown Menues: Klicken Sie auf die Zelle. Rechts neben der Zelle erscheint ein Button mit einem Pfeil nach unten. Klicken Sie auf diesen Button und das Menue wird angezeigt

Figure A.8.: Customer Intimacy Questionnaire: Work Environment (German)

B. Machine Learning Algorithms Settings

This appendix consists of five tables. The first four tables describe the considered parameters for configuring the machine learning algorithms used in this thesis and elaborated in chapter 7: the decision tree C4.5, the multilayer perceptron with backpropagation neural network, *k*-nearest neighbour, and the support vector machine. The description of the individual options is derived from the Weka documentation.[1] The exact list of parameter combinations tested in this project is available upon request from the author. Finally, table B.5 details the number of configurations tested for each machine learning algorithm and for each predicted variable.

---

[1] Further details are available at *http://www.cs.waikato.ac.nz/ml/weka/* (accessed on 1.12.2011).

Table B.1.: Configuration Settings of the Decision Tree C4.5

Option	Considered Values	Description
binarySplit	True / False	Whether data splits on nominal attributes are binary or not
confidenceFactor	From 0.1 to 0.8 with increments of 0.1	Degree of pruning of the tree
minNumObj	From 2 to 10 with increments of 1	Minimum number of objects per terminal leaf
reducedErrorPruning	True / False	Whether to use reduced error pruning instead of C4.5 error pruning or not
numFolds	From 2 to 5 with increments of 1	Amount of data used for reduced-error raising pruning (if reducedErrorPruning is set to true)
subTreeRaising	True / False	Whether to use the subtree raising operation or not during the pruning task
useLaplace	True / False	Whether to use the Laplace function when counting the the number of instances at a node
unpruned	True / False	Whether to perform the pruning task or not

Table B.2.: Configuration Settings of the *k*-nearest Neighbor Algorithm

Option	Considered Values	Description
KNN	From 1 to 10 with increments of 1	Number of neighbors to use
crossValidate	True / False	Use hold-one-out cross-validation to select the best k value
distanceWeighting	No Distance Weighting Weight by 1/distance Weight by 1 - distance	Determines the distance weighting method
meanSquared	True / False	Determines whether to use the mean squared error or the mean absolute error
NearestNeighbor SearchAlgorithm	LinearNNSearch BallTree CoverTree KDTree	The nearest neighbour search algorithm to use

Table B.3.: Configuration Settings of the Support Vector Machine Algorithm

Option	Considered Values	Description
buildLogisticModels	False	Whether to fit logistic models to the outputs
c	0.5 to 2.5 with increments of 0.1	Complexity Parameter C
Epsilon	$1.0.E^{-12}$	The epsilon for round-off error
Kernel	Polykernel Puk RBFKernel NormalizedPolyKernel	The Kernel to use
toleranceParameter	0.0010	The tolerance parameter

Table B.4.: Configuration Settings of the Multilayer Perceptron with Backpropagration

Option	Considered Values	Description
decay	True / False	Decreases the learning rate if set to true
learningRate	from 0.0 to 1.0	The amount the weights are updated
Momentum	from 0.0 to 1.0	Momentum applied to the weights during updating
NominalToBinary	True / False	Can improve performance if the data set contains binary attributes
Reset	True / False	
hiddenLayer	a	Determines the number of hidden layers automatically
autobuild	True	Adds and connects up hidden network automatically
trainingTime	1/500	The number of epochs to train through.
validationSetTime	0	No validation set will be used and instead the network will train for the specified number of epochs.

Table B.5.: Number of Tested Configurations of the Machine Learning Algorithms to Predict the Customer Intimacy Values

	Amount of Tested Configurations				
	C4.5	*k*-NN	SVM	NNBP	Total
Individual Level					
Knowledge High	96	36	41	71	244
Knowledge Very High	74	29	41	86	230
Relationship High	52	29	46	67	194
Relationship Very High	54	36	46	56	192
Organizational Level					
Knowledge High	27	43	46	57	173
Knowledge Very High	27	31	46	62	166
Relationship High	35	31	46	68	180
Relationship Very High	27	31	46	62	166
Total	392	266	358	529	1545

C. Acquired Customer Intimacy at the Individual Level

This appendix provides further details on the metrics calibration which is developed in chapter 7 to assess the acquired customer intimacy components at the individual level. Figure C.1 shows the Crombach's Alpha values of the summated scales *Knowledge* and *Relationship* at the individual level. Table C.2 details the achieved calibration results to predict the variable *Knowledge High* with the decision tree C4.5 algorithm. It can be observed that 52 models have been created and tested, the model number 40 obtaining the best results and being therefore chosen. Tables C.2, C.3, C.4 and C.5 detail the best calibration settings of the four considered machine learning algorithms to assess the variables *Knowledge High*, *Knowledge Very High*, *Relationship High*, and *Relationship Very High* at the individual level.

Case Processing Summary

		N	%
Cases	Valid	117	100.0
	Excluded[a]	0	.0
	Total	117	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.912	2

(a) Scale *Knowledge*

Case Processing Summary

		N	%
Cases	Valid	104	100.0
	Excluded[a]	0	.0
	Total	104	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.940	2

(b) Scale *Relationship*

Figure C.1.: Crombach's Alpha of the Scales *Knowledge* and *Relationship* at the Individual Level

Table C.1.: Prediction of the Variable *Knowledge High*: Detailed Performance Results of the Decision Tree C4.5

Model Number	Binary Split	Confidence Factor	MinNumObj	NumFolds	Reduced-Error-Pruning	SubTree-Raising	Unpruned	Use-Laplace	Success Rate (%)	Precision (%)	Recall (%)	F-Measure (%)	Kappa Statistic (%)
1	0	0.25	2	3	0	1	0	0	77.1	0.83	0.66	0.72	0.54
2	1	0.25	2	3	0	1	0	0	77.1	0.83	0.66	0.72	0.54
3	0	0.1	2	3	0	1	0	0	77.45	0.85	0.67	0.72	0.54
4	0	0.2	2	3	0	1	0	0	77.62	0.84	0.67	0.73	0.55
5	0	0.3	2	3	0	1	0	0	77.36	0.83	0.67	0.72	0.54
6	0	0.4	2	3	0	1	0	0	77.11	0.83	0.67	0.72	0.54
7	0	0.5	2	3	0	1	0	0	76.01	0.82	0.65	0.7	0.51
8	0	0.6	2	3	0	1	0	0	75.25	0.83	0.61	0.69	0.5
9	0	0.7	2	3	0	1	0	0	75.25	0.83	0.61	0.69	0.5
10	0	0.8	2	3	0	1	0	0	75.25	0.83	0.61	0.69	0.5
11	0	0.25	1	3	0	1	0	0	77.36	0.84	0.67	0.72	0.54
12	0	0.25	3	3	0	1	0	0	78.06	0.84	0.69	0.74	0.56
13	0	0.25	4	3	0	1	0	0	77.72	0.84	0.68	0.74	0.55
14	0	0.25	5	3	0	1	0	0	77.87	0.84	0.69	0.74	0.55
15	0	0.25	6	3	0	1	0	0	77.2	0.82	0.69	0.73	0.54
16	0	0.25	7	3	0	1	0	0	75.33	0.8	0.67	0.71	0.5
17	0	0.25	8	3	0	1	0	0	74.82	0.79	0.67	0.7	0.49
18	0	0.25	9	3	0	1	0	0	74.8	0.8	0.65	0.7	0.49
19	0	0.25	10	3	0	1	0	0	75.08	0.82	0.64	0.69	0.5
20	0	0.25	20	3	0	1	0	0	76.72	0.87	0.62	0.7	0.53
21	0	0.25	30	3	0	1	0	0	76.72	0.88	0.61	0.7	0.53
22	0	0.25	40	3	0	1	0	0	78.41	0.85	0.68	0.74	0.56
23	0	0.25	50	3	0	1	0	0	72.73	0.72	0.73	0.72	0.45
24	0	0.25	2	2	1	1	0	0	76.64	0.85	0.64	0.71	0.53
25	0	0.25	2	3	1	1	0	0	77.58	0.86	0.65	0.72	0.55
26	0	0.25	2	4	1	1	0	0	76.24	0.85	0.63	0.7	0.52
27	0	0.25	2	5	1	1	0	0	76.7	0.85	0.63	0.7	0.53
28	0	0.25	2	3	0	0	0	0	77.45	0.84	0.67	0.73	0.55
29	0	0.25	2	3	0	1	1	0	75.67	0.84	0.62	0.7	0.51
30	0	0.25	2	3	0	1	0	1	77.1	0.83	0.66	0.72	0.54

Continued

Model Number	Binary Split	Confidence Factor	MinNumObj	NumFolds	Reduced-Error-Pruning	SubTree-Raising	Unpruned	Use-Laplace	Success Rate (%)	Precision (%)	Recall (%)	F-Measure (%)	Kappa Statistic (%)
31	0	0.2	1	3	0	1	0	0	77.36	0.84	0.67	0.72	0.54
32	0	0.2	2	3	0	1	0	0	77.62	0.84	0.67	0.73	0.55
33	0	0.2	4	3	0	1	0	0	77.64	0.84	0.68	0.73	0.55
34	0	0.2	5	3	0	1	0	0	77.7	0.84	0.69	0.74	0.55
35	0	0.2	6	3	0	1	0	0	77.2	0.82	0.69	0.73	0.54
36	0	0.2	7	3	0	1	0	0	75.06	0.8	0.66	0.7	0.5
37	0	0.2	8	3	0	1	0	0	74.29	0.79	0.65	0.69	0.48
38	0	0.1	3	3	0	1	0	0	77.47	0.84	0.68	0.73	0.55
39	0	0.3	3	3	0	1	0	0	78.06	0.84	0.69	0.74	0.56
40	**0**	**0.4**	**3**	**3**	**0**	**1**	**0**	**0**	**78.39**	**0.84**	**0.7**	**0.75**	**0.56**
41	0	0.5	3	3	0	1	0	0	77.8	0.83	0.68	0.73	0.55
42	0	0.6	3	3	0	1	0	0	74.27	0.8	0.62	0.68	0.48
43	0	0.7	3	3	0	1	0	0	74.27	0.8	0.62	0.68	0.48
44	0	0.8	3	3	0	1	0	0	74.27	0.8	0.62	0.68	0.48
45	0	0.2	3	3	0	1	0	0	78.06	0.84	0.69	0.74	0.56
46	0	0.4	3	2	1	1	0	0	76.2	0.84	0.64	0.71	0.52
47	0	0.4	3	3	1	1	0	0	77.41	0.86	0.65	0.72	0.54
48	0	0.4	3	4	1	1	0	0	77.18	0.86	0.64	0.72	0.54
49	0	0.4	3	5	1	1	0	0	76.62	0.85	0.63	0.71	0.53
50	0	0.4	3	3	0	0	0	0	78.23	0.84	0.69	0.74	0.56
51	0	0.4	3	3	0	1	1	0	74.6	0.81	0.63	0.69	0.49
52	0	0.4	3	3	0	1	0	1	78.39	0.84	0.7	0.75	0.56

Table C.2.: Prediction of the Variable *Knowledge High* at the Individual Level: Best Configurations and Results

	Attribute	Value	Attribute	Value
Decision Tree C4.5	BinarySplit Minimum	False	Confidence Factor	0.4
	Number of Object per Leaf	3	Number of folds	N/A
	Reduced Error Pruning	False	SubTreeRaising	True
	Unpruned	False	UseLaplace	False
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	84.0	Recall(%)	70.0
	Success Rate (%)	78.4	F-Measure (%)	75.0
	Kappa Statistic (%)	56.0		
	Attribute	Value	Attribute	Value
***k*-nearest Neighbor**	kNN	9	distanceWeighting	No Distance Weighting
	meanSquared	False	NearestNeighbor SearchAlgorithm	Linear-NNSearch
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	83.0	Recall(%)	67.0
	Success Rate (%)	77.1	F-Measure (%)	72.0
	Kappa Statistic (%)	54.0		

Prediction of the Variable *Knowledge High* at the Individual Level: Best Configurations and Results (Continued)

	Attribute	Value	Attribute	Value
Support Vector Machine	Complexity c	1.5	Kernel	PolyKernel
	Epsilon	$1.0.E^{-12}$	Tolerance Parameter	0.001
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	87.0	Recall(%)	67.0
	Success Rate (%)	79.1	F-Measure (%)	74.0
	Kappa Statistic (%)	58.0		
	Attribute	Value	Attribute	Value
Neural Network	Autobuild	True	Decay	False
	hiddenLayer	a	LearningRate	0.1
	Momentum	0.17	Nominalto BinaryFilter	True
	Reset	True	ValidationSetSize	0
	Training Time	500	ValidationThreshold	20
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	87.0	Recall(%)	71.0
	Success Rate (%)	80.2	F-Measure (%)	76.0
	Kappa Statistic (%)	60.0		

Table C.3.: Prediction of the Variable *Knowledge Very High* at the Individual Level: Best Configurations and Results

	Attribute	Value	Attribute	Value
Decision Tree C4.5	BinarySplit Minimum	False	Confidence Factor	0.4
	Number of Object per Leaf	6	Number of folds	N/A
	Reduced Error Pruning	False	SubTreeRaising	True
	Unpruned	False	UseLaplace	False
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	72.0	Recall(%)	55.0
	Success Rate (%)	79.4	F-Measure (%)	59.0
	Kappa Statistic (%)	46.0		
	Attribute	Value	Attribute	Value
k-nearest Neighbor	kNN	3	distanceWeighting	No Distance Weighting
	meanSquared	False	NearestNeighbor SearchAlgorithm	Linear-NNSearch
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	83.0	Recall(%)	57.0
	Success Rate (%)	83.5	F-Measure (%)	64.0
	Kappa Statistic (%)	55.0		

Prediction of the Variable *Knowledge Very High* at the Individual Level: Best Configurations and Results (Continued)

	Attribute	Value	Attribute	Value
Support Vector Machine	Complexity c	1.7	Kernel	Normalized PolyKernel
	Epsilon	$1.0.E^{-12}$	Tolerance Parameter	0.001
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	71.0	Recall(%)	62.0
	Success Rate (%)	80.3	F-Measure (%)	63.0
	Kappa Statistic (%)	50.0		
	Attribute	Value	Attribute	Value
Neural Network	Autobuild hiddenLayer	True a	Decay LearningRate	False 0.2
	Momentum	0.1	Nominalto BinaryFilter	True
	Reset Training Time	True 100	ValidationSetSize ValidationThreshold	0 20
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	61.0	Recall(%)	60.0
	Success Rate (%)	76.5	F-Measure (%)	58.0
	Kappa Statistic (%)	42.0		

Table C.4.: Prediction of the Variable *Relationship High* at the Individual Level: Best Configurations and Results

	Attribute	Value	Attribute	Value
Decision Tree C4.5	BinarySplit Minimum	False	Confidence Factor	0.20
	Number of Object per Leaf	2	Number of folds	N/A
	Reduced Error Pruning	False	SubTreeRaising	True
	Unpruned	False	UseLaplace	False
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	80.0	Recall(%)	59.0
	Success Rate (%)	67.0	F-Measure (%)	65.0
	Kappa Statistic (%)	35.0		
	Attribute	Value	Attribute	Value
k-nearest Neighbor	kNN	6	distanceWeighting	No Distance Weighting
	meanSquared	False	NearestNeighbor SearchAlgorithm	Linear-NNSearch
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	80.0	Recall(%)	75.0
	Success Rate (%)	73.3	F-Measure (%)	76.0
	Kappa Statistic (%)	45.0		

Prediction of the Variable *Relationship High* at the Individual Level: Best Configurations and Results (Continued)

	Attribute	Value	Attribute	Value
Support Vector Machine	Complexity c	1.2	Kernel	PolyKernel
	Epsilon	$1.0.E^{-12}$	Tolerance Parameter	0.001
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	86.0	Recall(%)	69.0
	Success Rate (%)	75.4	F-Measure (%)	75.0
	Kappa Statistic (%)	51.0		
	Attribute	Value	Attribute	Value
Neural Network	Autobuild	True	Decay	False
	hiddenLayer	a	LearningRate	0.2
	Momentum	0.1	Nominalto BinaryFilter	True
	Reset	True	ValidationSetSize	0
	Training Time	50	ValidationThreshold	20
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	79.0	Recall(%)	72.0
	Success Rate (%)	70.9	F-Measure (%)	73.0
	Kappa Statistic (%)	41.0		

Table C.5.: Prediction of the Variable *Relationship Very High* at the Individual Level: Best Configurations and Results

	Attribute	Value	Attribute	Value
Decision Tree C4.5	BinarySplit Minimum	False	Confidence Factor	0.1
	Number of Object per Leaf	7	Number of folds	N/A
	Reduced Error Pruning	False	SubTreeRaising	True
	Unpruned	False	UseLaplace	False
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	75.0	Recall(%)	52.0
	Success Rate (%)	81.1	F-Measure (%)	58.0
	Kappa Statistic (%)	48.0		
	Attribute	Value	Attribute	Value
k-nearest Neighbor	kNN	5	distanceWeighting	Weight by 1 - distance
	meanSquared	False	NearestNeighbor SearchAlgorithm	Linear-NNSearch
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	66.0	Recall(%)	55.0
	Success Rate (%)	77.8	F-Measure (%)	57.0
	Kappa Statistic (%)	43.0		

Prediction of the Variable *Relationship Very High* at the Individual Level: Best Configurations and Results (Continued)

	Attribute	Value	Attribute	Value
Support Vector Machine	Complexity c	1	Kernel	PolyKernel
	Epsilon	$1.0.E^{-12}$	Tolerance Parameter	0.001
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	65.0	Recall(%)	52.0
	Success Rate (%)	77.4	F-Measure (%)	54.0
	Kappa Statistic (%)	41.0		
Neural Network	Attribute	Value	Attribute	Value
	Autobuild	True	Decay	False
	hiddenLayer	a	LearningRate	0.3
	Momentum	0.2	Nominalto BinaryFilter	True
	Reset	True	ValidationSetSize	0
	Training Time	2000	ValidationThreshold	20
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	50.0	Recall(%)	51.0
	Success Rate (%)	74.9	F-Measure (%)	47.0
	Kappa Statistic (%)	34.0		

D. Acquired Customer Intimacy at the Organizational Level

This appendix complements chapter 7 and provides further details on the performed calibration to predict the acquired customer intimacy components at the organizational level. Figure D.1 outlines the Crombach's Alpha values of the summated scales *Knowledge* and *Relationship* at the organizational level. Tables D.1, D.2, D.3 and D.4 present the settings of four considered machine learning algorithms which achieved the best prediction of the variables *Knowledge High*, *Knowledge Very High*, *Relationship High*, and *Relationship Very High* at the organizational level.

Case Processing Summary

		N	%
Cases	Valid	77	100.0
	Excluded[a]	0	.0
	Total	77	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.911	3

(a) Scale *Knowledge*

Case Processing Summary

		N	%
Cases	Valid	73	100.0
	Excluded[a]	0	.0
	Total	73	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.891	3

(b) Scale *Relationship*
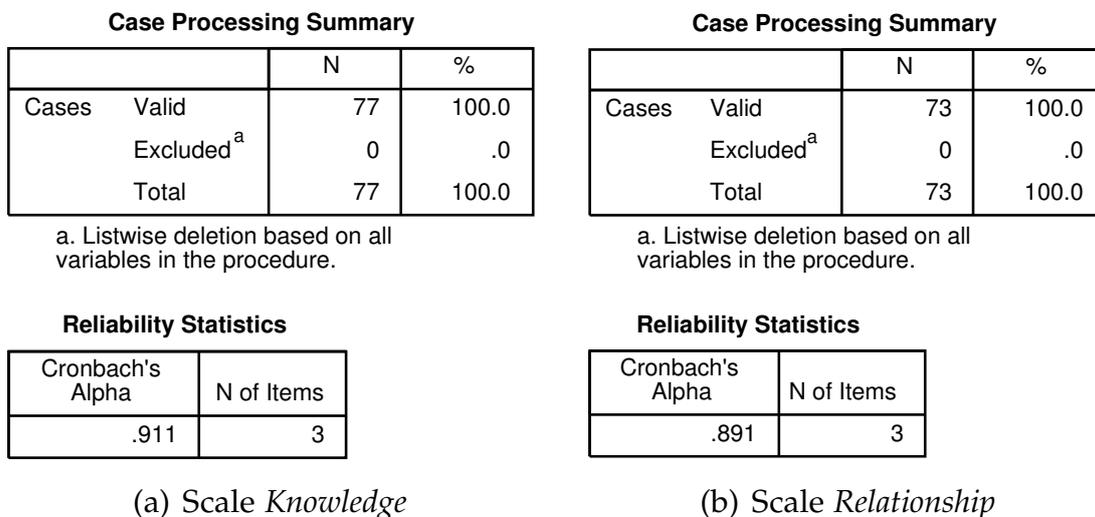
Figure D.1.: Crombach's Alpha of the Scales *Knowledge* and *Relationship* at the Organizational Level

Table D.1.: Prediction of the Variable *Knowledge High* at the Organizational Level: Best Configurations and Results

	Attribute	Value	Attribute	Value
Decision Tree C4.5	BinarySplit Minimum	False	Confidence Factor	0.2
	Number of Object per Leaf	4	Number of folds	N/A
	Reduced Error Pruning	False	SubTreeRaising	True
	Unpruned	False	UseLaplace	False
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	73.0	Recall(%)	65.0
	Success Rate (%)	79.6	F-Measure (%)	66.0
	Kappa Statistic (%)	53.0		
	Attribute	Value	Attribute	Value
k-nearest Neighbor	kNN	10	distanceWeighting	No Distance Weighting
	meanSquared	False	NearestNeighbor SearchAlgorithm	Linear-NNSearch
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	73.0	Recall(%)	51.0
	Success Rate (%)	78.9	F-Measure (%)	58.0
	Kappa Statistic (%)	47.0		

Prediction of the Variable *Knowledge High* at the Organizational Level: Best Configurations and Results (Continued)

	Attribute	Value	Attribute	Value
Support Vector Machine	Complexity c	2.5	Kernel	RBFKernel
	Epsilon	$1.0.E^{-12}$	Tolerance Parameter	0.001
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	75.0	Recall(%)	81.0
	Success Rate (%)	82.1	F-Measure (%)	75.0
	Kappa Statistic (%)	61.0		
	Attribute	Value	Attribute	Value
Neural Network	Autobuild	True	Decay	False
	hiddenLayer	a	LearningRate	0.2
	Momentum	0.2	Nominalto BinaryFilter	True
	Reset	True	ValidationSetSize	0
	Training Time	150	ValidationThreshold	20
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	50.0	Recall(%)	54.0
	Success Rate (%)	68.3	F-Measure (%)	48.0
	Kappa Statistic (%)	29.0		

Table D.2.: Prediction of the Variable *Knowledge Very High* at the Organizational Level: Best Configurations and Results

	Attribute	Value	Attribute	Value
Decision Tree C4.5	BinarySplit Minimum	False	Confidence Factor	0.2
	Number of Object per Leaf	2	Number of folds	N/A
	Reduced Error Pruning	False	SubTreeRaising	True
	Unpruned	False	UseLaplace	False
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	41.0	Recall(%)	45.0
	Success Rate (%)	84.6	F-Measure (%)	40.0
	Kappa Statistic (%)	35.0		
	Attribute	Value	Attribute	Value
k-nearest Neighbor	kNN	4	distanceWeighting	No Distance Weighting
	meanSquared	False	NearestNeighbor SearchAlgorithm	Linear-NNSearch
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	42.0	Recall(%)	39.0
	Success Rate (%)	87.6	F-Measure (%)	38.0
	Kappa Statistic (%)	36.0		

Prediction of the Variable *Knowledge Very High* at the Organizational Level: Best Configurations and Results (Continued)

	Attribute	Value	Attribute	Value
Support Vector Machine	Complexity c	0.4	Kernel	PolyKernel
	Epsilon	$1.0.E^{-12}$	Tolerance Parameter	0.001
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	32.0	Recall(%)	35.0
	Success Rate (%)	80.1	F-Measure (%)	32.0
	Kappa Statistic (%)	23.0		

	Attribute	Value	Attribute	Value
Neural Network	Autobuild	True	Decay	False
	hiddenLayer	a	LearningRate	0.2
	Momentum	0.2	Nominalto BinaryFilter	True
	Reset	True	ValidationSetSize	0
	Training Time	400	ValidationThreshold	20
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	14.0	Recall(%)	19.0
	Success Rate (%)	80.1	F-Measure (%)	14.0
	Kappa Statistic (%)	10.0		

Table D.3.: Prediction of the Variable *Relationship High* at the Organizational Level: Best Configurations and Results

	Attribute	Value	Attribute	Value
Decision Tree C4.5	BinarySplit Minimum	False	Confidence Factor	0.6
	Number of Object per Leaf	9	Number of folds	N/A
	Reduced Error Pruning	False	SubTreeRaising	True
	Unpruned	False	UseLaplace	False
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	55.0	Recall(%)	55.0
	Success Rate (%)	56.0	F-Measure (%)	52.0
	Kappa Statistic (%)	11.0		
	Attribute	Value	Attribute	Value
k-nearest Neighbor	kNN	4	distanceWeighting	No Distance Weighting
	meanSquared	False	NearestNeighbor SearchAlgorithm	Linear-NNSearch
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	53.0	Recall(%)	61.0
	Success Rate (%)	56.5	F-Measure (%)	55.0
	Kappa Statistic (%)	13.0		

Prediction of the Variable *Relationship High* at the Organizational Level: Best Configurations and Results (Continued)

	Attribute	Value	Attribute	Value
Support Vector Machine	Complexity c	2.5	Kernel	RBFKernel
	Epsilon	$1.0.E^{-12}$	Tolerance Parameter	0.001
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	64.0	Recall(%)	74.0
	Success Rate (%)	66.8	F-Measure (%)	66.0
	Kappa Statistic (%)	33.0		
	Attribute	Value	Attribute	Value
Neural Network	Autobuild	True	Decay	False
	hiddenLayer	a	LearningRate	0.1
	Momentum	0.2	Nominalto BinaryFilter	True
	Reset	True	ValidationSetSize	0
	Training Time	9	ValidationThreshold	20
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	54.0	Recall(%)	67.0
	Success Rate (%)	58.6	F-Measure (%)	58.0
	Kappa Statistic (%)	17.0		

Table D.4.: Prediction of the Variable *Relationship Very High* at the Organizational Level: Best Configurations and Results

	Attribute	Value	Attribute	Value
Decision Tree C4.5	BinarySplit Minimum	False	Confidence Factor	0.3
	Number of Object per Leaf	0.3	Number of folds	N/A
	Reduced Error Pruning	False	SubTreeRaising	False
	Unpruned	False	UseLaplace	False
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	48.0	Recall(%)	49.0
	Success Rate (%)	78.2	F-Measure (%)	44.0
	Kappa Statistic (%)	33.0		
	Attribute	Value	Attribute	Value
k-nearest Neighbor	kNN	4	distanceWeighting	No Distance Learning
	meanSquared	False	NearestNeighbor SearchAlgorithm	Linear-NNSearch
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	41.0	Recall(%)	29.0
	Success Rate (%)	79.6	F-Measure (%)	33.0
	Kappa Statistic (%)	24.0		

Prediction of the Variable *Relationship Very High* at the Organizational Level: Best Configurations and Results (Continued)

	Attribute	Value	Attribute	Value
Support Vector Machine	Complexity c	0.4	Kernel	PolyKernel
	Epsilon	$1.0.E^{-12}$	Tolerance Parameter	0.001
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	32.0	Recall(%)	35.0
	Success Rate (%)	80.1	F-Measure (%)	32.0
	Kappa Statistic (%)	23.0		
Neural Network	Attribute	Value	Attribute	Value
	Autobuild	True	Decay	False
	hiddenLayer	a	LearningRate	0.2
	Momentum	0.2	Nominalto BinaryFilter	True
	Reset	True	ValidationSetSize	0
	Training Time	200	ValidationThreshold	20
	Performance Indicator	Value	Performance Indicator	Value
	Precision (%)	18.0	Recall(%)	16.0
	Success Rate (%)	77.6	F-Measure (%)	16.0
	Kappa Statistic (%)	11.0		

E. CI Analytics Implementation

This appendix complements chapter 6 and provides further details on the implementation of the software *CI Analytics*. Section E.1 develops the services which have been conceived and implemented to calculate the acquired customer intimacy metrics. Section E.2 elaborates on the services designed to calculate the leveraged customer intimacy metrics. Subsequently, section E.3 introduces the software *CI Graph* which is the first prototype of the software *CI Analytics* and which was realized together with Thomas Herzig. Finally, section E.4 presents the questionnaire used to performed the survey on the business benefits of the software *CI Analytics* and details the survey participants profiles.

E.1. CI Services for Calculating the Acquired Customer Intimacy Metrics

This appendix elaborates on the CI Services realized to calculate the acquired customer intimacy metrics which are presented in section 6.2.4.

As detailed in table E.1, the services focusing on the individual level of analysis return a graph representing the social network formed by the employees of the provider and of the customer. These employees are represented by nodes on the graph and the customer intimacy metrics values are indicated by the weights of the graph edges. The social network graph returned by these CI Services is presented in the XML format DyNetML which has been specifically conceived for the representation of social networks (Tsvetovat *et al.*, 2004). The services calculate the customer intimacy metrics upon the data available in the customer interaction time fact table and take the seven following parameters as input:

- *CustomerOrgRef* determines the customer organization for which the customer intimacy metrics are calculated

- *StartDate* and *EndDate* determine the beginning and the end of the considered time frame.

- *SegmentSize* specifies the length of each segment in the time period and determines, therefore, the precision of the analysis.

- *InteractionDurationThreshold*, *InteractionQuantityThreshold* and *WeightedInteractionQuantityThreshold* allow to further calibrate the calculation of the customer intimacy metrics, as detailed in section 5.2.2.1

The services realized to calculate the acquired customer intimacy at the organization level of analysis return the value of the customer intimacy metric between a specific provider employee and the considered customer organization. Similarly to the services created to calculate the acquired customer intimacy at the individual level, these services use the data available in the customer interaction time fact table in order to calculate the metrics. In addition to the input parameters defined for the services performing the calculation at the individual level, the services calculating the acquired customer intimacy at the organizational level also take the reference to a specific provider employee as input parameter, as depicted in table E.1.

Table E.1.: CI Services For Calculating the Acquired Customer Intimacy Metrics: Technical Details

CI Services at the Individual Level

Name	Volume Service, WVolume Service, Intensity Service, WIntensity Service, Frequency Service, Duration Service, NumberEpisodes Service, Mode Service
Input Parameters	CustomerOrgRef (String), StartDate (Integer), EndDate (Integer), SegmentSize (Integer), InteractionDurationThreshold (Integer), InteractionQuantityThreshold (Integer), WeightedInteractionQuantityThreshold (Integer)
Output Value	Social Network Graph (DyNetML Format)
Fact Table	Customer Interaction Time Fact Table
Description	Eight CI services provide the functionality to calculate the acquired customer intimacy metric at the individual level.

CI Services at the Organizational Level

Name	Org Volume Service, Org WVolume Service, Org Intensity Service, Org WIntensity Service, Org Frequency Service, Org Duration Service, Org NumberEpisodes Service, Org Mode Service
Input Parameters	CustomerOrgRef (String), ProviderEmployeeRef (String), StartDate (Integer), EndDate (Integer), SegmentSize (Integer), InteractionDurationThreshold (Integer), InteractionQuantityThreshold (Integer), WeightedInteractionQuantityThreshold
Output Value	MetricValue (Double)
Fact Table	Customer Interaction Time Fact Table
Description	Eight organizational CI services provide the functionality to calculate the acquired customer intimacy metrics at the organizational level.

E.2. CI Services for Calculating the Leveraged Customer Intimacy Metrics

This appendix elaborates on the CI Services realized to calculate the seven leveraged customer intimacy metrics which are presented in section 6.2.4. Table E.2 describes each of these seven services and provides information on their inputs and outputs.

Table E.2.: CI Services for the Leveraged Customer Intimacy Metrics

Service Name	
Customization Revenue Ratio Service	
Component	Customization
Metric	*Customization Revenue Ratio*
Input Parameters	CustomerOrgRef (String), StartDate (Integer), EndDate (Integer)
Output Parameter	Ratio value (Double) comprised between 0 and 1
Fact Table	Customer Value Return
Description	This service provides the functionality to calculate the customer intimacy metric *Customization Revenue Ratio* to assess the degree of customization of the solution provided to the customer. To perform the calculation, this service only uses the monetary revenues derived from business objects of type *Invoice Line Item* and excludes non monetary returns such as customer suggestions.
Customer Purchase Frequency Ratio Service	
Component	Customer Loyalty
Metric	*Customer Purchase Frequency Ratio*
Input Parameters	CustomerOrgRef(String), StartDate (Integer), EndDate (Integer)
Output Parameter	Frequency value (Double) comprised between 0 and 1
Fact Table	Customer Value Return (only monetary revenues)

| Description | This service provides the functionality to calculate the customer intimacy metric *Customer Purchase Frequency Ratio* which has been established as an indicator of the customer loyalty. |

CrossSelling Revenue Ratio Service

Component	Cross-Selling
Metric	*Cross-Selling Revenue Ratio*
Input Parameters	CustomerOrgRef (String), StartDate (Integer), EndDate (Integer)
Output Parameter	Ratio value (Double) comprised between 0 and 1
Fact Table	Customer Value Return
Description	This service enables the calculation of the customer intimacy metric *Cross-Selling Revenue Ratio*. Similarly to the *Customization Revenue Ratio Service*, the *CrossSelling Revenue Ratio Service* only considers the monetary revenues recorded in the Customer Value Return fact table. This service identifies the source of the revenues such as product and service reference numbers that the customer purchased for the first time within the time period. It then calculates the *Cross-Selling Revenue Ratio* as the ratio between the revenues generated from these sources and the total revenues generated in the considered time period with the customer.

CrossSelling Diversity Ratio Service

Component	Cross-Selling
Metric	*Cross-Selling Diversity Ratio*
Input Parameters	CustomerOrgRef(String), StartDate (Integer), EndDate (Integer)
Output Parameter	Ratio value (Double) comprised between 0 and 1
Fact Table	Customer Value Return
Description	This service provides the functionality to calculate the customer intimacy metric *Cross-Selling Diversity Ratio* upon the monetary revenues recorded in the fact table Customer Value Return.

Customer Participation Quantity Service

Component	Customer Participation
Metric	*Customer Participation Quantity*
Input Parameters	CustomerOrgRef (String), StartDate (Integer), EndDate (Integer)
Output Parameter	Participation Quantity (Double)
Fact Table	Customer Value Return (excluding monetary revenues)
Description	This service calculates the metric *Customer Participation Quantity* upon the data available in the Customer Value Return fact table as the total number of suggestions submitted by the customer in the considered time frame.

Customization Participation Ratio Service

Metric	*Customer Participation Ratio*
Input Parameters	CustomerOrgRef(String), StartDate (Integer), EndDate (Integer)
Output Parameter	Ratio value (Double) comprised between 0 and 1
Fact Table	Customer Value Return
Description	This service provides the functionality to calculate the customer intimacy metric *Customer Participation Ratio*. It accesses the data contained in the Customer Value Return fact table, calculates the number of suggestions performed by the customer during a certain time frame, and divides this value by the revenues generated with this customer during the same time frame.

Transaction Effectiveness Ratio Service

Component	Transaction Cost Reduction
Metric	*Transaction Effectiveness Ratio*
Input Parameters	CustomerOrgRef (String), StartDate (Integer), EndDate (Integer)
Output Parameter	Ratio value (Double) comprised between 0 and 1

Fact Table	Customer Value Return
Description	This service provides the functionality to calculate the customer intimacy metric *Transaction Effectiveness Ratio*. It accesses the data contained in the three fact tables Customer Value Return, Customer Activity Time and Customer Interaction Time. The total interaction time and activity time that occurred within the considered time period are divided by the total revenues generated with the customer during the same time period.

E.3. *CI Graph*: A First Prototype of *CI Analytics*

This section introduces the first prototype called *CI Graph* of the software *CI Analytics*. *CI Graph* has been conceived and implemented to generate the data set on which the calibration of the customer intimacy metrics presented in chapter 7 has been performed.

While the software *CI Analytics* in its current version includes both the acquired and leveraged customer intimacy metrics and adheres to business intelligence application standards, the objective of the application *CI Graph* was to prove the feasibility of the measurement of the acquired customer intimacy metrics and of the representation of these metrics by means of a social network graph. Therefore, the software *CI Graph* provides the functionality to measure and visualize the eight *customer interaction time* based acquired customer intimacy metrics which are presented in chapter 5 at both the individual and organizational levels. *CI Graph* is also capable of calculating the centrality metrics developed in section 5.2.3.

Figure E.2 illustrates the architecture of the software *CI Graph*. This architecture consists of multiple modules developed in the C# language. In order to access the data contained in the database of the application CAS genesisWorld, the *Data Access* module of *CI Graph* does not use an ETL process directly accessing the database. Instead, *CI Graph* requests and receives the data through the CAS genesisWorld server, using an API of the CAS genesisWorld server. Thus, the performance impact on CAS genesisWorld is significantly higher with *CI Graph* than with *CI Analytics*.
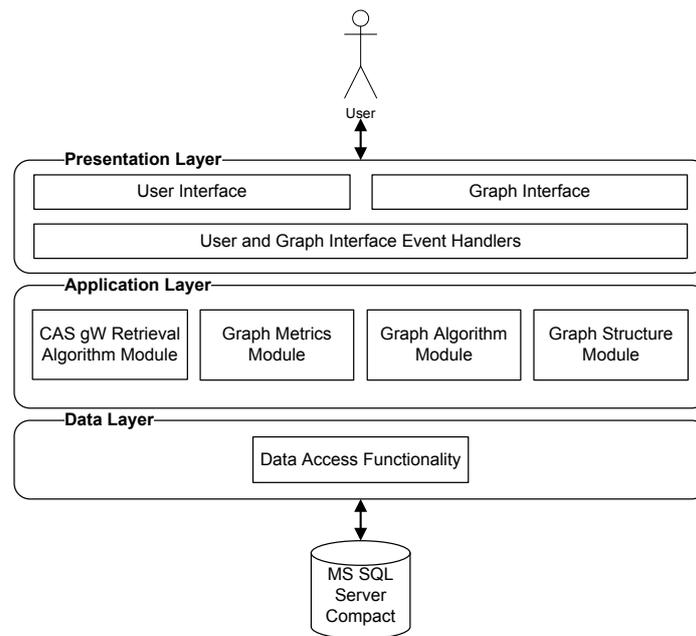
Figure E.2.: *CI Graph*: Architecture Overview

In order to calculate the customer intimacy metrics, the *CAS gWRetrieval* module retrieves the required interaction data from CAS GenesisWorld, calculates upon predefined calibration parameters (time period T, segment size d, interaction quantity threshold b and weighted interaction quantity threshold wb) the *customer interaction time* and *weighted customer interaction time* for each segment in the considered time period. The acquired customer intimacy metrics are subsequently calculated upon this data by the functions implemented in the *Graph Metrics* module and stored into a table contained in a database. The data in this table has been used to perform the calibration of the customer intimacy metrics presented in chapter 7. In order to represent the customer intimacy metrics in the form of a social network, the *Graph Structure* module uses the data contained in the previously populated table and creates the graph representation of the social network upon customer intimacy metric selected by the user. The *Graph Algorithm* module can finally be used in order to calculate the network centrality metrics upon this graph. Further details on the architecture and implementation of the software *CI Graph* are available upon request from the author.

Figure E.3 illustrates the calibration panel of the software *CI Graph*. The user enters its credentials and the location of the CAS genesis-World server in order to connect to CAS genesisWorld server and to retrieve the required interaction data. The user subsequently enters the name of a customer organization and a date used to specify the considered time period: the calculation is performed for the year preceding the date entered in this panel. Finally, the user clicks on the button StartAnalysis in order to begin the metric calculation process.

The graph panel of the software *CI Graph* allows a visualization of the acquired customer intimacy by means of a social network, as depicted in figure E.4. In this diagram, the rectangles in the top row represent the customer employees and those in the bottom row the provider employees. The values of the acquired customer intimacy metrics are indicated by the weights of the edges on the graph. The interface provides the ability to select between the three time periods, namely 3 months, 12 months or all-time, as well as to select a metric to be visualized on the edges and a layout for the network representation. Clicking on the button CalculateGraphMetrics initiates the process of calculating the network centrality metrics at the organizational level.
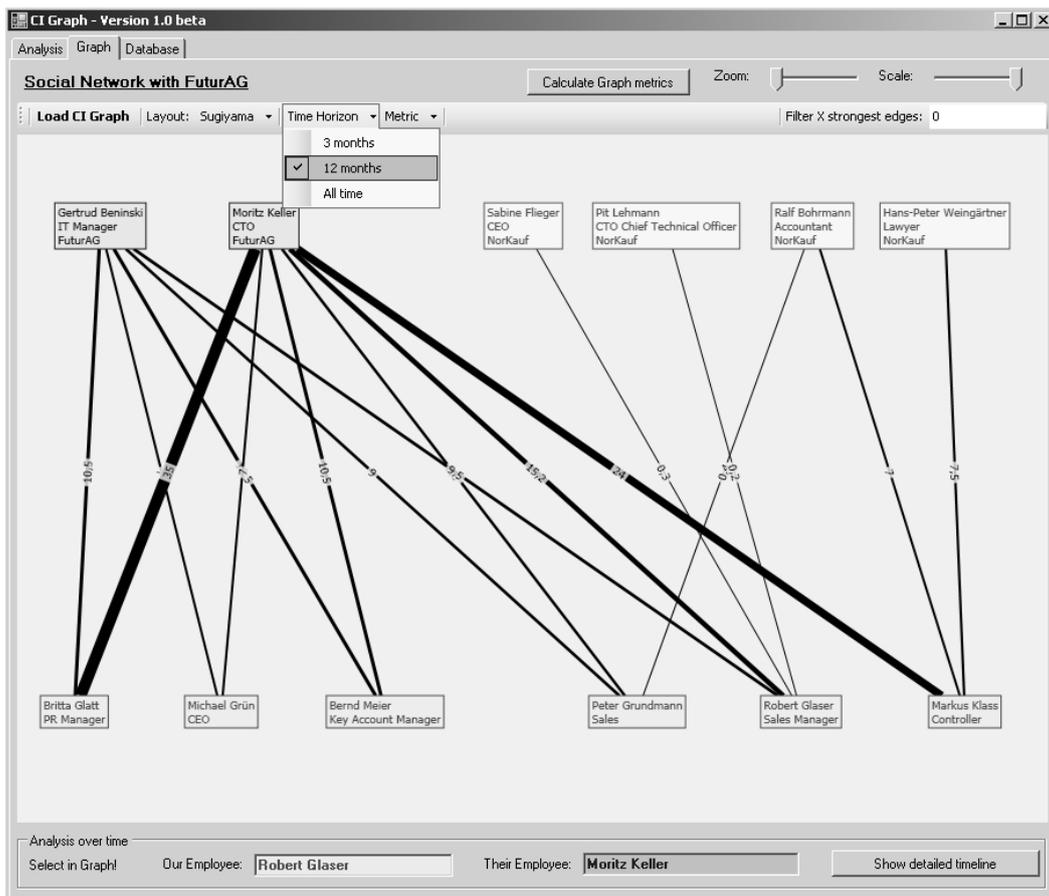
Figure E.3.: *CI Graph*: Calibration Panel

Figure E.4.: *CI Graph*: Visualization Panel

E.4. Business Benefits Analysis

This section complements the business benefits survey developed in section 6.3.2. Figures E.5, E.6 and E.7 represent the questionnaire designed to assess the business benefits of the software *CI Analytics*. Figure E.8 provides further details on the survey participants with regard to their role in the organization and to their interactions with customers.

Thank you for participating in this interview!

It will take you less than 10 minutes to complete it.

Please return your answers by Friday September 3rd to
Thomas Herzig:
thomas.herzig@student.kit.edu

What do I have to do?
1. Read and understand the context
2. Answer questions by ticking the appropriate box.

What will happen with my answers?
Your answers help us evaluate the business benefits of our research and our prototype. All
answers and responses will be handled confidentially and anonymously at all times.

What is this interview about?
We are conducting research on an automatic analysis of interaction between companies
(from interaction data contained in an enterprise IT system). The results are used in a
customer relationship management application that shows a social network between
employees of a service provider and employees of their customers. The objective of this
prototype is to help the service provider employees answer questions like:

- "We are starting a new project with a team from CustomerXY, does someone from my
  company already know them?"
- "Have we cultivated our relationship with the customer during the last months?"

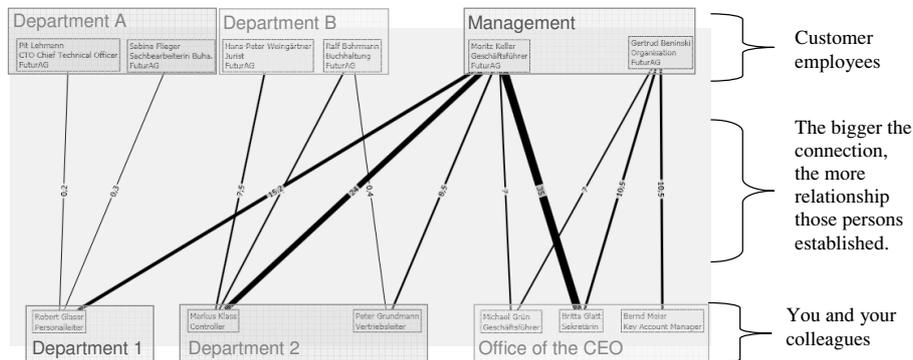The Relationship Network Overview (simplified example)

Figure E.5.: *CI Analytics*: Business Benefits Questionnaire (1/3)

Please provide some information about your activities

Question1: What is your **role** inside your company?

☐   ☐   ☐   ☐   ☐   ☐
Services   Sales   Marketing   Development   Management   Other

Question2: How many **customer companies** have you worked with during the last year?

☐   ☐   ☐
Less than 3   Between 3 an 10   More than 10

Question3: How many **employees** from customers did you have contact with during the last year?

☐   ☐   ☐
Less than 10   Between 10 and 50   More than 50

Question4: How much of your **time** did you spend working with customers during the last year?

☐   ☐   ☐
Less than 20%   Between 20 and 50%   More than 50%

Now please consider customers with whom you have worked with during the past year and imagine you had the relationship network overview presented above available

Question5: I would use this overview to identify colleagues who have knowledge about the customer organization (strategy, processes, organization, behaviour, etc.)

☐   ☐   ☐   ☐   ☐
Strongly agree   Agree   No opinion   Disagree   Strongly disagree

Question6: I would use this overview to identify colleagues who have established relationships with the customer employees

☐   ☐   ☐   ☐   ☐
Strongly agree   Agree   No opinion   Disagree   Strongly disagree

Question7: This relationship network overview would help us share knowledge about the customer inside our organization

☐   ☐   ☐   ☐   ☐
Strongly agree   Agree   No opinion   Disagree   Strongly disagree

Question8: This relationship network overview would help us coordinate our activities towards the customer and to be seen as one team by the customer

☐   ☐   ☐   ☐   ☐
Strongly agree   Agree   No opinion   Disagree   Strongly disagree

Question9: Analyzing the evolution of this relationship network overview over time would help us monitor the relationship with the customer (e.g. "Have we cultivated the relationship with the customer during the last months?")

☐   ☐   ☐   ☐   ☐
Strongly agree   Agree   No opinion   Disagree   Strongly disagree

Question10: Together with other indicators such as sales results, this information would help us compare the performance achieved with different customers and would help us in our choice to invest in one or the other customer.

☐   ☐   ☐   ☐   ☐
Strongly agree   Agree   No opinion   Disagree   Strongly disagree

Figure E.6.: *CI Analytics*: Business Benefits Questionnaire (2/3)

You are almost done, two last questions:

Question11: I think such a visualization would be useful in our company

☐   ☐   ☐   ☐   ☐
Strongly agree   Agree   No opinion   Disagree   Strongly disagree
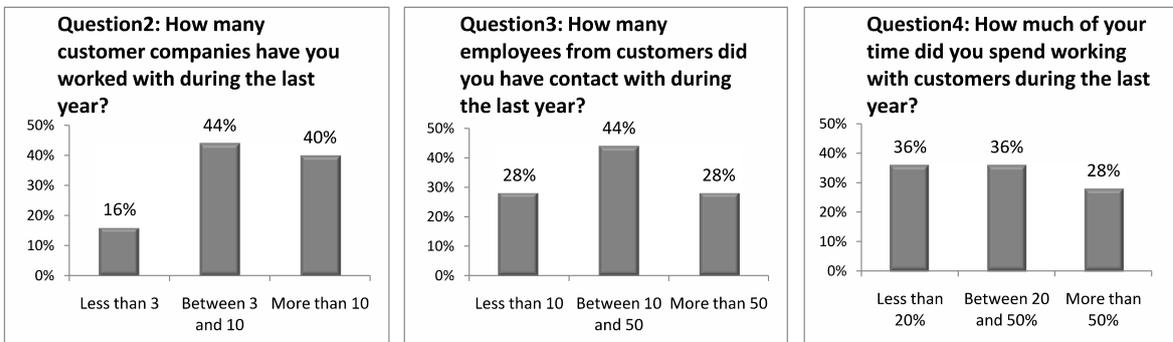
Question12: I would have privacy concerns if this type of information was made available in my company

☐   ☐   ☐   ☐   ☐
Yes, strong concerns   Yes, I have concerns   No opinion   No concerns   No, absolutely no concerns

Figure E.7.: *CI Analytics*: Business Benefits Questionnaire (3/3)
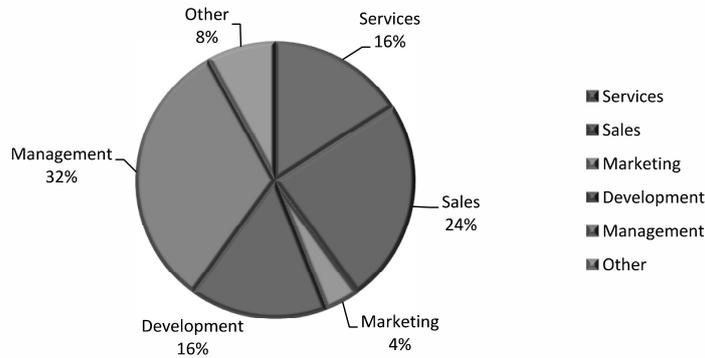
(a) Question 1

(b) Question 2,3,4

Figure E.8.: Business Benefits Survey: Participants Profiles

François Habryn
CUSTOMER INTIMACY ANALYTICS

The ability to capture customer needs and to tailor the provided solutions accordingly, also defined as customer intimacy, has become a significant success factor in the B2B space – in particular for increasingly "servitizing" businesses. However, many organizations struggle with measuring and proactively managing the degree of customer intimacy established with their customers. The work presented in this book aims to remedy this issue by providing a solution to the assessment and monitoring of the key aspects of a customer intimacy strategy. It leverages business analytics and social network analysis technology in order to provide an accurate, real-time, and easily implementable assessment of customer intimacy, thereby effectively complementing existing customer relationship management systems.

This book proposes a solid, innovative and clearly written contribution that should be of interest to all business and IT leaders facing the challenges of customer intimacy (Prof. Dr. Gerhard Satzger).

François Habryn is a senior research associate at the Karlsruhe Service Research Institute. He gained several years of experience in IT consulting with IBM and holds a Ph.D. in economics from the Karlsruhe Institute of Technology. François Habryn graduated from the University of Technology of Compiègne in France with a Master's degree in computer science and from the Ecole Supérieure de Commerce de Paris (ESCP-Europe) with a Master's degree in European business.