



Big Data und KI im Fußball

Themenkurzprofil Nr. 72 | Simone Ehrenberg-Silies | Juni 2024

Big Data und künstliche Intelligenz (KI) prägen inzwischen das Fußballspiel. Die massenhafte Erfassung von Datenströmen im Profibereich bildet die Grundlage für die Datenrevolution des Fußballs. Während des Spiels werden Positionsdaten sämtlicher Spieler/innen erfasst. Hinzu kommen Ereignisdaten, die Sachverhalte, wie beispielsweise Pässe, Schüsse, Zweikämpfe, Auswechslungen, festhalten. In Verbindung mit KI verändern Fußballdaten schon heute das Scouting, die Spiel-, Spieler/innen- und Gegneranalyse, die Art und Weise, wie Trainer/innen während des Spiels intervenieren, Livekommentierungen und wie Zuschauer/innen das Spiel erleben. So groß das Anwendungspotenzial der KI-gestützten Datenanalyse im Fußball ist, so voraussetzungsreich ist auch die Datennutzung. Beispielsweise kann die Datenerhebung nicht in jedem Fall automatisiert erfolgen und Daten müssen teilweise bereinigt oder validiert werden. Zudem setzt die Erfassung der Daten Technologien wie kamerabasierte Trackingsysteme oder ein geschultes Team zur Liveerfassung voraus. Überdies sind die Datenverfügbarkeit – Positionsdaten liegen überwiegend für die TOP-Ligen der Männer vor – und ihre Vergleichbarkeit wie bei den Ereignisdaten unterschiedlicher Anbieter eingeschränkt. Eine zusätzliche Herausforderung ist der fehlende Konsens darüber, was in Bezug auf Spieler/innen oder Mannschaften überhaupt für den Spielerfolg relevante Leistungsindikatoren sind und wie diese einheitlich definiert und damit modelliert werden können. Außerdem setzt die Nutzung der Daten ein hohes Maß an Daten- und Fußballexpertise voraus. Wenngleich die Vermessung des Fußballs suggeriert, dass dieser vorhersehbarer würde, ist anzunehmen, dass auch im Zeitalter von KI und Big Data der Zufall im Fußball weiterhin eine entscheidende Rolle für das Spielergebnis innehaben wird. Deshalb gilt heute wie 2006: „So ist Fußball. Manchmal gewinnt der Bessere!“ (Lukas Podolski).

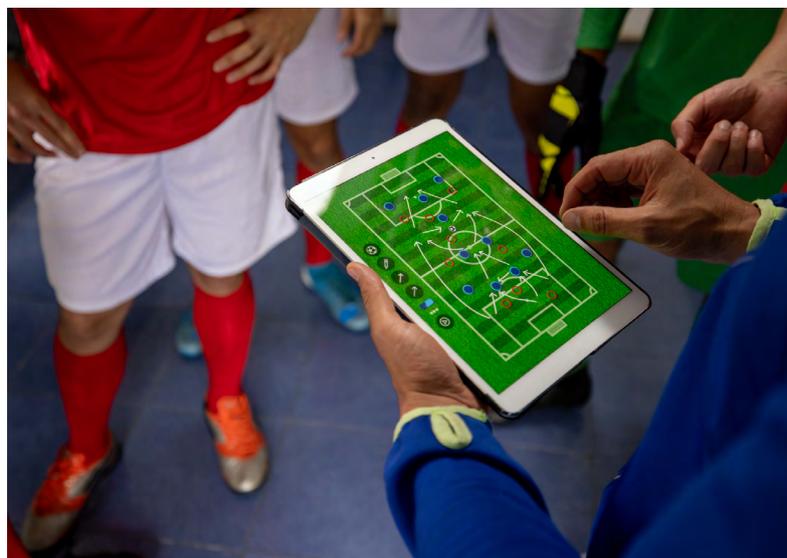
Hintergrund und Entwicklungsstand

Geschichte der Datenrevolution

Die „Datenrevolution“ (Biermann 2018, S.64) des Fußballs hat ihren Ursprung gewissermaßen im Baseball: In den USA begann der Baseballfan Bill James in den 1970er Jahren, Statistiken zu Aktionen und Spielzügen im Baseball anzufertigen, und prägte später den Begriff „SABRmetrics“, der sich von der 1971 gegründeten Society for American Baseball Research (SABR) ableitete. Er und weitere Mitglieder von SABR versuchten, auf Basis von erhobenen Daten aus zahlreichen Baseballspielen mathematisch zu analysieren, welche Parameter Erfolge und Misserfolge von Mannschaften beeinflussten. Da Baseball im Vergleich zum Fußball aus vielen statischen Spielsituationen besteht, lässt es sich statistisch wesentlich leichter erfassen als das fluide Fußballspiel (Huber 2023). Anfang der 2000er Jahre machte sich Billy Beane, ehemaliger Scout und General Manager der Oakland Athletics in der Major League Baseball, die Datenrevolution zunutze und führte sein Team mehrere Jahre in die Play-offs, obwohl es mit den Budgets anderer Mannschaften nicht mithalten konnte (Biermann 2018, S.64 f.). Während für die Aufstellung eines Kaders bis zu diesem Zeitpunkt üblicherweise die Einschätzungen und Beobachtungen des Trainerteams zu einzelnen Spielern ausschlaggebend waren, wählte Billy Beane Spieler auf Grundlage bestimmter Leistungsdaten und Statistiken für die Oakland Athletics aus. So gelang es ihm, talentierte, bislang unbekannte Spieler mit einem relativ kleinen Budget zu verpflichten (Diez 2022). Einem größeren (Sport-)Publikum wurde die Strategie Billy Beanes durch das 2004 von Michael Lewis erschienene Buch „Moneyball: The Art of Winning an Unfair Game“ bekannt. Seit dieser Veröffentlichung erreichte die Datenrevolution in den Vereinigten Staaten immer mehr Teams in der Major League Baseball und weitere Ballsportarten wie American Football, Basketball und bedingt auch Eishockey (Biermann 2018, S.64 f.).

Bis die Datenrevolution im europäischen Fußball ankommen sollte, dauerte es noch beinahe ein Jahrzehnt (Hummel/Karon o.J.). Hier machten der dänische Verein FC Midtjylland und der FC Liverpool aus der englischen Premier League als Pioniere im Bereich Datenanalyse und künstlicher Intelligenz von sich reden. Der Physiker Matthew Bentham kaufte 2014 den damaligen dänischen Provinzclub; zuvor war er mit Datenanalysetools auf dem Wettmarkt sehr erfolgreich gewesen. Aufgrund der positiven Erfahrung seines Besitzers mit Datenanalyse setzte der Verein fortan mathematische Modelle und Algorithmen zur Spiel-, Gegner- und Spieleranalyse ein und nutzte sie auch zum Scouting. 2015 wurde der FC Midtjylland erstmals dänischer Meister und konnte sogar in der Europa League einen Sieg gegen Manchester United erringen (Biermann 2018, S.73; Brüggemann 2022, S.127). Die Geschichte des FC Liverpool weist einige Parallelen zur Geschichte des Einzugs der Datenanalyse beim FC Midtjylland auf: Der Gründer des amerikanischen Sportvermarktungskonzerns Fenway Sports Group, dem auch der FC Liverpool angehört, gelangte als Rohstoffhändler an der Börse zu Reichtum. Das Geheimnis seines Erfolgs war die Entwicklung eines Algorithmus, mit dem er Schwankungen beim Kurs von Soja prognostizieren konnte. Möglicherweise waren die positiven Erfahrungen mit Datenanalyse im Rohstoffhandel ausschlaggebend für seinen Versuch, mit dem gleichen Ansatz im Fußball zu reüssieren. So stellte der FC Liverpool 2012 den Physiker Ian Graham ein und baute eine eigene Datenabteilung mit fünf Datenanalysten auf (Stand 2021). Mithilfe von Scouting- und Trackingdaten gelang es dem FC Liverpool wie auch dem FC Midtjylland und den Oakland Athletics im Baseball, Spieler mit großem Potenzial wie Virgil van Dijk (Abwehr), Philippe Coutinho (Sturm) und Mohamed Salah (Sturm) zu verpflichten und diese zu Spielern mit einem zwischenzeitlich enormen Marktwert weiterzuentwickeln (Hummel/Karon o.J.). Virgil van Dijk wechselte Anfang 2018 für eine Ablösesumme von 84,65 Mio. Euro von Southampton zum FC Liverpool (damaliger Marktwert 30 Mio. Euro) und erreichte 2019 einen Marktwert von 100 Mio. Euro (Transfermarkt o.J.a). Philippe Coutinho wechselte Anfang 2013 für 13 Mio. Euro (damaliger Marktwert 9 Mio. Euro) von Inter Mailand zum FC Liverpool und konnte Anfang 2018 für 135 Mio. Euro an den FC Barcelona verkauft werden (Transfermarkt o.J.b). Für Mohamed Salah zahlte der FC Liverpool dem AS Rom Mitte 2017 eine Ablösesumme von 42 Mio. Euro (damaliger Marktwert 35 Mio. Euro). 2018 erreichte der Spieler einen Marktwert von 150 Mio. Euro (Transfermarkt o.J.c).

Im Gegensatz zur Bundesliga sind Datenexpert/innen in der Premier League mittlerweile weit verbreitet. Dabei gilt der FC Liverpool zusammen mit den Vereinen Brighton und Brentford nach wie vor als Vorreiter. Auch die italienische Serie A (Vorreiter AC Mailand), die französische Ligue 1 (Vorreiter FC Toulouse) und die US-amerikanisch Major League Soccer sind nach Einschätzung Ian Grahams beim Aufbau eigener Datenexpertise und -analyseabteilungen weiter fortgeschritten als die Bundesligavereine (Hofmann 2024).



Dabei werden Daten – auch in der Bundesliga – inzwischen massenhaft erhoben (Huber 2023). Für die 1. und 2. Bundesliga erhebt die DFL-Tochtergesellschaft Sportec Solutions die offiziellen Spieldaten. Diese umfassen neben Positions- und Ereignisdaten auch Stammdaten und Spielinformationen und werden den Bundesligavereinen, Lizenznehmern und Partnern der Deutschen Fußball Liga (DFL) zur Verfügung gestellt (DFL o.J.). Seit der Saison 2022/2023 erhebt Sportec Solutions zudem die Daten im DFB-Pokal und seit der Saison 2023/2024 in der 3. Liga und der Google Pixel Frauen-Bundesliga (DFB 2023).

Positionsdaten

Positionsdaten geben Aufschluss darüber, zu welchem Zeitpunkt sich welcher Spieler wo auf dem Platz aufhält und wo sich der Ball befindet (Schramm 2023). Bisher werden sie überwiegend in den TOP-Ligen im Männerfußball systematisch erhoben (Memmert 2022, S.7). Sie werden in Form von x-y-Koordinaten ausgegeben (Diez 2022; Memmert 2022, S.10). Um sie zu erfassen, setzt Sportec Solutions in der Bundesliga in Abhängigkeit von der Spielstätte 16, 18 oder 20 Kameras ein (DFL o.J.). Der Ball hat einen eigenen Sensor (Schramm 2023). Mittels kamerabasiertem Tracking können die jeweiligen Positionen der Spieler auf dem Platz ermittelt werden. Die Spielerpositionen werden über Bilderkennung- und Interpolationsverfahren aus den unterschiedlichen Kameraperspektiven abgeleitet (Raabe 2022, S.238). Insgesamt fallen 3,6 Mio. Positionsdatenpunkte pro Spiel an (DFL o.J.). Alternativ zu kamerabasierten Trackingsystemen kann die Spielerposition auch mittels tragbarer Sensorik, beispielsweise über das Globale Positionsbestimmungssystem (GPS), erfasst werden. Die durch GPS generierten Daten sind für physische Belastungsmessungen geeignet, z. B. die absolvierte Laufstrecke, für taktische Analysen jedoch zu ungenau. Die vergleichsweise beste Datengenauigkeit kann durch Local- Positioning-Systeme (LPS) erreicht werden. Hierfür werden Antennen rund um das Spielfeld aufgestellt. Diese fangen Radiofrequenzsignale von den Wearables der Spieler auf, sodass mithilfe von Triangulation deren Position bestimmt werden kann (Lutz et al. 2020; Raabe 2022,

S.238). Der Aufbau solcher Systeme ist jedoch aufwendig (Raabe 2022, S.238), was vermutlich dazu beiträgt, dass sie zur Erhebung von Positionsdaten bisher kaum verbreitet sind. Dies gilt ebenso für GPS-Sensoren (Schramm 2023).

Positionsdaten sind mit einer leichten Verzögerung quasi in Echtzeit verfügbar. Waren sie vor 10 bis 15 Jahren noch sehr fehlerhaft (ein Spieler scheint auf der Tribüne zu sitzen und taucht kurze Zeit später auf dem Rasen auf), hat sich die Datenqualität inzwischen deutlich verbessert. Das Tracking hat aber auch Tücken, etwa wenn Spieler übereinander fallen. In diesen Fällen ist nicht eindeutig detektierbar, welcher Spieler aufsteht und zu welcher Mannschaft er gehört. Durch automatisierte Prüfroutinen werden die Daten bereits während des Spiels bereinigt. Die Datenbereinigung ist allerdings bisher nicht vollautomatisiert möglich, sodass hierfür menschliche Spielbeobachter/innen eingesetzt werden müssen (Schramm 2023).

Aus den Positionsdaten lassen sich Informationen wie beispielsweise Laufgeschwindigkeiten und Entfernungen direkt ableiten (Memmert 2022, S.7). Weitergehende Auswertungen setzen jedoch „fortgeschrittene Konzepte und Algorithmen“ voraus (Raabe 2022, S.239). Grundsätzlich sind die Möglichkeiten der Modellierungen auf Basis von Positionsdaten enorm. Allerdings ist derzeit noch weitgehend unklar, wie spieltaktisch relevante Leistungsindikatoren (Key Performance Indicators) aussehen könnten. Dies hängt u.a. damit zusammen, dass Fußballtaktiken wie Pressingstrategien unterschiedlich definiert werden. Operationalisierbare Definitionen sind jedoch Voraussetzung für jede quantitative Modellkonstruktion (Raabe 2022, S.239). Leistungsindikatoren, die sich auf Basis von Positionsdaten bisher bewährt haben, sind der Voronoi-Ansatz zur Bemessung der Raumkontrolle und die Anzahl der durch Pässe überspielten bzw. noch zu überspielenden Gegenspieler. In Feldstudien und Experimenten konnte gezeigt werden, dass diese beiden Leistungsindikatoren eine signifikante Korrelation mit dem Spielausgang bzw. mit der Saisonleistung haben. Sie werden deshalb situativ auch zum Vergleich zwischen Spielsystemen und zur Bewertung einzelner Spieler herangezogen. Ein tiefer gehendes Verständnis der Leistungsindikatoren im Kontext des Fußballspiels ist für ihre Interpretation allerdings unabdingbar: So eignet sich die Raumkontrolle nur eingeschränkt für die Bewertung individualtaktischer Leistungen. Die Werte, die einzelne Spieler beim Leistungsindikator Anzahl überspielter Gegenspieler erzielen können, hängen zudem von der jeweiligen Spielerposition ab: Innenverteidiger und defensive Mittelfeldspieler erreichen hier prinzipbedingt bessere Ergebnisse als Stürmer (Raabe 2022, S.240 ff.).

Ereignisdaten

Ereignisdaten umfassen ballspezifische (z.B. Pässe, Schüsse, Zweikämpfe), systemisch-technische (z.B. Auswechslungen, Fouls) und zeitliche Ereignisse im Fußballspiel, etwa den Beginn der Spielzeit. Es wird erfasst, wann die jeweiligen

Ereignisse stattgefunden haben, wo dies auf dem Spielfeld geschehen ist, welche Spieler/innen involviert waren und was das Ergebnis des Ereignisses ist. Unter Umständen wird auch die Spielrichtung erfasst. Durch die Analyse der Zeit- und Koordinatenattribute von Ereignissen können Erkenntnisse über das „Was, Wann und Wo“ geliefert werden (Garnica Caparrós 2022, S.174 u. 178). Ereignisdaten haben die Funktion, einzelne Ereignisse oder die Abfolge von Ereignissen aufzuzeigen. Die Erfassung von Ereignisdaten ist weiter verbreitet als die Erfassung von Positionsdaten. So liegen Ereignisdaten auch für die meisten semiprofessionellen Ligen weltweit vor (Memmert 2022, S.7f.). Mittlerweile bieten auch vermehrt Drittanbieter Ereignisdaten für Jugendligen oder kleinere Wettbewerbe an (Garnica Caparrós 2022, S.179). Die Daten werden während des Spiels erhoben oder aus Videomaterial gewonnen (Memmert 2022, S.7f.). Für Bundesligaspiele setzt Sportec Solutions vier Mitarbeiter/innen zur Erfassung der Ereignisdaten ein. Ein/e Mitarbeiter/in im Stadion übermittelt einem/einer Mitarbeiter/in im Sportec Solutions Collection Centre den Spielverlauf in einer codierten Sprache. Diese/r pflegt die Ereignisse in Echtzeit in ein System ein. Die zwei weiteren Personen validieren die eingegebenen Ereignisse auf Basis von Livevideoausschnitten (DFL o.J.).

An dieser Vorgehensweise wird bereits ein wesentlicher Nachteil von Ereignisdaten deutlich. In sie fließt Fußballwissen zu einzelnen Aktionen ein, das nicht so einfach zu standardisieren ist. Infolgedessen sind Ereignisdaten verschiedener Anbieter wie Opta Sports, Wyscout oder StatsBomb nicht per se vergleichbar (Garnica Caparrós 2022, S.176). Ein weiterer Nachteil von Ereignisdaten besteht darin, dass sie spielrelevante Situationen ohne Ballbesitz nicht abbilden und nicht den gesamten Spielzusammenhang erfassen können (Raabe 2022, S.238).

Ereignisdaten lassen sich z.B. zur Erzeugung von Boxscores verwenden. Boxscores umfassen bestimmte Ereignisse wie Tore, Torschüsse und Assists (Torvorlagen) bezogen auf einzelne Spieler/innen. Auf ihrer Grundlage lässt sich beispielsweise beantworten, wie viele Pässe ein/e Spieler/in spielte und wie viele davon prozentual Offensivpässe waren. Ein Nachteil von Boxscores ist, dass sie sämtliche, in den Ereignisdaten eigentlich enthaltene Kontextdaten (z.B. was vor dem Foul passierte) ignorieren (Garnica Caparrós 2022, S.177). Nichtsdestotrotz können auf Basis von Boxscores Spieler/innen miteinander verglichen werden. Teilweise geschieht dies in Verbindung mit anderen Erkenntnissen aus Ereignisdaten, z.B. zum/zur dominanten Spieler/in mit hoher Passaktivität durch die Visualisierung von Passnetzwerken (Clemente et al. 2015, nach Garnica Caparrós 2022, S.178).

Da sich auf der Basis von Ereignisdaten zudem einzelne Spielsequenzen im Hinblick auf Erfolg und Misserfolg bewerten lassen, bildeten sie die Grundlage für die Entwicklung des Expected-Goal(xG)-Modells und korrespondierender Modelle wie Expected Assists (xA) oder Expected Threat (xT)

(Garnica Caparrós 2022, S.177 ff.). Mithilfe von Ereignisdaten lässt sich auch berechnen, inwieweit einzelne Aktionen von Spieler/innen einen Effekt auf das Ergebnis haben, also die Wahrscheinlichkeit eines erfolgreichen Torschusses der eigenen Mannschaft erhöhen und der gegnerischen Mannschaft reduzieren. Das hierfür eingesetzte Modell wird „Expected-Value-of-Possession“ genannt (Brunnschneider/Hahn 2022, S.186 u. 188).

Gesellschaftliche und politische Relevanz

Obwohl es noch einige ungeklärte Fragen hinsichtlich der Datenqualität, -vergleichbarkeit und -verwendbarkeit im Fußball gibt, ergeben sich durch die Datenanalyse mit KI vielfältige Anwendungsmöglichkeiten, die u. a. das Scouting, die Spiel-, Gegner- und Spieleranalyse, das Coaching, die Fußballberichterstattung sowie das Fußballerlebnis für Zuschauer/innen deutlich verändern könnten.

Scouting

Zurzeit setzt die überwiegende Mehrheit der Profivereine das Scouting noch konventionell durch Spielebeobachtung um: Lediglich rund 10 % der Profivereine weltweit nutzen Datenscouting (Klingelhöfer 2023). Datenscouting wird jedoch voraussichtlich mit der Verbreitung von Datenerfassung in allen möglichen Ligen der Welt deutlich zunehmen. Es wird prognostiziert, dass Datenscouting bald auch in die Jugendbereiche U17 und U19 Einzug halten wird. Durch den Preisrückgang bei Kamerasystemen ist zudem auch ein Vordringen in den Amateurbereich vorstellbar. In Zukunft könnte es möglich sein, Spieler/innen aus den Ligen weltweit auf der Basis von Leistungsindikatoren und Metriken, die für den jeweiligen Club und sein Team von besonderer Relevanz sind, miteinander zu vergleichen (Diez 2022; Klingelhöfer 2023).

Zahlreiche Dienstleister bieten bereits heute datenbasierte Unterstützung für das Scouting an. Global Soccer Network (GSN) bewertet beispielsweise eine halbe Million Fußballspieler/innen weltweit. Der Anbieter erfasst Basisdaten der Spieler/innen wie Größe, Gewicht und Alter und sammelt zudem Positions-, Ereignis- und Leistungsdaten. Daraus lässt sich u. a. die Torgefährlichkeit eines Teams berechnen oder die Auswirkung einer einzelnen Spieleraktion für das Team analysieren. Diese Datensets lassen sich weitgehend automatisch erstellen (Schramm 2023).

Faktoren wie Mentalität, Kreativität oder die Kommunikation von Spieler/innen untereinander, die ebenfalls die Leistung der/des Einzelnen und des Teams beeinflussen, werden gegenwärtig und voraussichtlich auch in absehbarer Zukunft noch manuell generiert werden müssen. Zwar sind sie nicht objektiv, jedoch häufig zumindest intersubjektiv: Menschliche Scouts bewerten auf Basis von 130 vorgegebenen Attributen unabhängig voneinander taktische, technische, mentale und physische Kompetenzen von Spieler/innen. Dabei geht es z. B. um die Beweglichkeit in engen Situationen, die

Herangehensweise bei flachen und hohen Bällen, die Stärke des linken und rechten Fußes und das Verhalten gegenüber Schiedsrichter/innen, Mit- und Gegenspieler/innen. Liegen erst einmal ausreichend historische Daten zu Spieler/innen vor, könnte in Zukunft mithilfe eines Algorithmus die Frage beantwortet werden, wie sich Spieler/innen mit ähnlichen Fähigkeiten in der Vergangenheit entwickelt haben. Daraus könnten ggf. sogar Schlüsse für die Zusammenstellung des Kaders gezogen werden (Schramm 2023).

Mit der voraussichtlich zunehmenden Verbreitung der Datenerfassung von Spieler/innen werden auch Datenschutzfragen virulenter. Der Erfassung und Vermarktung der Spieldaten durch die Sportec Solutions haben alle Vereine und Spieler/innen zugestimmt. Die Antwort auf die Frage, was passieren würde, wenn Spieler/innen ihr Einverständnis nicht erteilen würden, ist offen. Es ist nicht unwahrscheinlich, dass sie im Profibereich keinen Vertrag mehr erhalten würden (Schramm 2023). Bei Anbietern, die ihre Daten aus Fernseh Bildern extrahieren, ist dem/der einzelnen Spieler/in zudem eventuell gar nicht bekannt, dass ein Datenset zu ihm/ihr angeboten und verkauft wird.

In Großbritannien ist es bereits zu einem Rechtsstreit zwischen Spieler/innen und Firmen gekommen (Schramm 2023). Das Project „Red Card“ richtet sich gegen die Verwertung von Spielerdaten durch Datenfirmen, ohne dass das Einverständnis der Spieler/innen vorliegt und sie dafür finanziell entschädigt werden. Planten zunächst 850 Fußballspieler/innen, die Klage gegen die Datenindustrie einzureichen, haben sich dieser inzwischen weitere Profisportler/innen aus den Sportarten Cricket und Rugby angeschlossen, sodass nun 1.400 Personen die Klage unterstützen (Hartley 2021; Seoudy 2023). Als kritischer Reflex zum gegenwärtigen Umgang mit Daten im Profifußball lassen sich auch die von der Fédération Internationale des Associations de Footballeurs Professionnels (FIFPRO), der weltweiten Organisation von Profifußballer/innen, veröffentlichte Charter of Player Data Rights (Charta der Datenrechte von Spieler/innen; FIFPRO 2022) sowie auch die Ankündigung, eine eigene zentralisierte, umfassende Datenplattform für Spielerdaten aufzubauen, interpretieren. Bei der Datenplattform der FIFPRO (2023) sollen Spieler/innen ihre Daten selbst verwalten können. Auch die deutsche Spielergewerkschaft Vereinigung der Vertragsfußballspieler e. V. (VDV o.J.), die im Übrigen kein Mitglied der FIFPRO ist, setzt sich für eine bessere Achtung der Persönlichkeitsrechte und der Privatsphäre von Spieler/innen, insbesondere im Hinblick auf den Schutz von Gesundheitsdaten, ein.

Zwar liegen in dem absehbaren Vordringen der Datenerfassung in den Jugend- und Amateurbereich viele Chancen, beispielsweise hinsichtlich der Datenvermarktung kleinerer Vereine in Zusammenarbeit mit Technologieanbietern (Klingelhöfer 2023) und der Erhöhung der Entdeckungswahrscheinlichkeit von talentierten Spieler/innen. Problematisch könnte es jedoch werden, wenn sich Jugendliche oder noch



nicht professionelle Spieler/innen gezwungen fühlen würden, der Datenerfassung zuzustimmen, da sie anderenfalls beim Datenscouting nicht berücksichtigt werden können und sich deshalb ihre Karriereperspektiven verschlechtern könnten.

Nachteilig für Spieler/innen könnten sich auch Messfehler (Schramm 2023) in Datenbanken erweisen, die ihre Leistungen schlechter erscheinen lassen, als sie eigentlich sind. Gegebenenfalls könnte hier das Auskunftsrecht gemäß Datenschutzgrundverordnung genutzt werden, um entsprechende Fehler aufzudecken.

Spielanalyse und Coaching

Im Unterschied zur auf Big Data und KI basierenden Gegnervorbereitung und Spielbeobachtung sind die Datengrundlage, die Objektivität und die Analysetiefe der konventionellen Spielanalyse notwendigerweise begrenzt. Die Spielanalysten/innen bzw. das Trainerteam können zur Gegnervorbereitung aus Zeitgründen nicht unendlich viel Videomaterial sichten. Sie müssen einzelne relevante Szenen für die Spielvorbereitung aus dem Videomaterial manuell zusammenstellen und können die Anpassung der Spielstrategie nur aus ihren Beobachtungen ableiten. Langzeitanalysen, beispielsweise über mehrere Spiele und Spielsaisons hinweg, im Hinblick auf Spieltaktiken und ihre Effizienz sowie objektive Hinweise in Echtzeit an das Trainerteam während des Spiels, wenn beispielsweise die Ballrückeroberung nach Gegenpressing zu lange dauert, sind mit einer konventionellen Herangehensweise der Spielvorbereitung und des Coachings nicht möglich (Anzer et al. 2022, S.261).

Mit KI verändert sich dies: Durch den Einsatz maschinellen Lernens (ML) können aus Positions- und Ereignisdaten taktische Muster im Fußball, wie das Gegenpressing, erkannt werden. Unter Gegenpressing versteht man das „aktive Ausüben von Raum- oder Zeitdruck auf den Gegner durch einen oder mehrere Spieler einer Mannschaft nach einem Ballverlust“ (Anzer et al. 2022, S.260). Durch diese taktische Vorgehensweise soll der Ball schnellstmöglich zurückerobert werden oder das gegnerische Team zu Befreiungsschlägen oder Rückpässen veranlasst werden (Anzer et al. 2022, S.260). Durch maschinelle Taktikererkennung können zur Gegnervorbereitung Videoplaylists mit spezifischen taktischen Situationen sowie objektive Trendanalysen zum Spielgeschehen automatisch erzeugt werden und somit die Entwicklung einer Spielstrategie unterstützen. Ebenso können ML-Algorithmen dazu verwendet werden, bei bestimmten Entwicklungen im Spielgeschehen Warnsignale an das Trainerteam zu senden (Anzer et al. 2022, S.260 f.).

Die Entwicklung weiterer KI-basierter Unterstützungsmöglichkeiten für Trainer/innen ist wahrscheinlich. So entstand aus der Zusammenarbeit zwischen dem FC Liverpool und Google Deepmind das Tool „TacticAI“, das Trainer/innen dabei assistieren soll, bei Eckbällen eine ideale Spielerkonstellation zu finden, je nachdem, ob der Eckball vom eigenen Team ausgeführt oder verteidigt wird (Engelien 2024).

Aus der Perspektive von Clubs und Trainer/innen ist ebenfalls das KI-Potenzial im Bereich der Verletzungsprävention relevant. Auf Grundlage historischer Daten könnten Algorithmen Muster für Verletzungsrisiken erkennen. Ausfälle von Schlüs-

selspieler/innen könnten so minimiert bzw. perspektivisch auch Kadergrößen reduziert werden. Die Forschung steckt hier allerdings noch in den Kinderschuhen und ist wegen der Verwendung von persönlichen Gesundheitsdaten von Spieler/innen (Diez 2022) unter dem Gesichtspunkt des Datenschutzes kritisch zu betrachten.

Veränderung des Zuschauererlebnisses

Seit 2020 stellt die DFL (Bundesliga 2021) in Zusammenarbeit mit Amazon Web Services (AWS) den Inhabern von Medienrechten die „Bundesliga Match Facts powered by AWS“ zur Verfügung. In der Bundesligasaison 2023/2024 wurden insgesamt 15 „Bundesliga Match Facts“ angeboten, die von „Expected Goals“, also den erwarteten Toren, und der Schussgeschwindigkeit über die Balleroberungsdauer bis zur Leistung unter Gegnerdruck und der Realformation, also dem tatsächlich praktizierten Spielsystem, reichen (Bundesliga o.J.a). Zuschauer/innen von Fußballspielen haben dadurch schon während des Spiels die Möglichkeit, die Leistungen der Teams auf Basis der angebotenen Match Facts zu vergleichen, und können auf diese Weise die eventuell als langatmig empfundenen Spielphasen ohne Torgefahr durch einen Blick aufs Smartphone überbrücken (WDR 5 2023).

Einige dieser Bundesliga Match Facts werden im Folgenden kurz vorgestellt.

- Der xG- oder auch xGoals-Wert gibt die Wahrscheinlichkeit an, mit einem Schuss auf das Tor tatsächlich einen Treffer zu erzielen. Das Modell wurde auf Basis von historischen Daten zu 40.000 Torschüssen entwickelt. Die Basis für die Berechnung des xG-Werts für jeden Torschuss in Echtzeit bilden die von Sportec Solutions erhobenen Positionsdaten, aus denen sich die für den xG-Wert relevanten Messgrößen wie die Distanz und der Winkel zum Tor, die Laufgeschwindigkeit des/der Spieler/in, die Zahl von zu überwindenden Gegenspieler/innen bis zum Tor und die Torabdeckung durch den/die Keeper/in ableiten lassen (Bundesliga 2020a). Seit der Bundesligasaison 2021/2022 spiegelt der xG-Wert überdies wider, unter welchen Bedingungen aufs Tor gezielt wird – etwa, ob es zu Einschränkungen bei der Ausführung gekommen ist. Der Wert für den einzelnen Torschuss liegt immer zwischen 0 und 1. Mit den xG-Werten lassen sich nicht nur die Torwahrscheinlichkeiten einzelner Torschüsse, sondern auch die individuellen Leistungen von Spieler/innen und Mannschaften unabhängig von den tatsächlich in einem Match erzielten Toren bewerten. Um den xG-Wert einer Mannschaft bzw. eine/r Spieler/in zu berechnen, werden die einzelnen xG-Werte ihrer Torchancen summiert. Hat ein/e Spieler/in beispielsweise einen individuellen xG-Wert von 14,7, aber 20 Tore erzielt, könnten hieraus Schlüsse auf seine/ihre Effizienz gezogen werden. Ein Nachteil des xG-Werts zur individuellen Leistungsbeurteilung besteht jedoch darin, dass bei seiner Berechnung kein Unterschied gemacht wird, ob ein/e Abwehrspieler/in oder eine Stürmer/in aufs Tor zielt (Bundesliga 2021).

- Die Schussgeschwindigkeit wird auf Grundlage der Positionsdaten des Balls bei Schüssen auf das Tor als Höchstwert der Geschwindigkeit auf seiner Flugbahn ermittelt. Die Schussgeschwindigkeit hat Einfluss darauf, ob ein/e Torwart/in einen Ball halten kann oder nicht (Bundesliga 2023a).
- Mit der Balleroberungsdauer wird gemessen, wie viel Zeit ein Team nach Ballverlust bis zur Rückeroberung des Balles benötigt. Die Balleroberungsdauer kann Hinweise auf die generelle oder situative taktische Ausrichtung eines Teams geben und anzeigen, ob es eher hoch presst oder sich in die eigene Hälfte zurückzieht (Bundesliga 2023b).
- Die Leistung unter Gegnerdruck beschreibt, in wie vielen Situationen sich Spieler/innen, die im Ballbesitz sind, durch Pässe, Dribblings, gezogene Fouls oder einen Torabschluss aus Spielsituationen befreien können, in denen sie vom gegnerischen Team unter Druck gesetzt werden (Bundesliga 2022).
- Die Realformation beschreibt auf Basis der 3,6 Mio. während eines Spiels erhobenen Datenpunkte, welches Spielsystem die einzelnen Teams während der gesamten Spielzeit, in Phasen mit und ohne Ball sowie in den letzten 10 Spielminuten auf dem Platz tatsächlich umsetzen (Bundesliga o.J.b). Auf dieser Grundlage wird ein Vergleich mit der nominellen Mannschaftsaufstellung möglich. Wird also tatsächlich durchgängig eine 4-4-2-Formation über die gesamte Spieldauer erkennbar oder zeigt sich zwischenzeitlich eine Dreierkette? Wie verhält sich das Team in den letzten 10 Spielminuten bei einer Führung, einem Rückstand oder bei Gleichstand? Taktikanpassungen je nach Spielverlauf können wichtige Hinweise auf das Verhalten gegnerischer Teams geben (Bundesliga 2020b).

Die Bundesliga Match Facts unterstützen auch Kommentator/innen bei ihrer (Live-)Berichterstattung. Den Journalist/innen wird basierend auf den Daten des „Bundesliga Data Hubs“ zusätzlich das Tool „Data Story Finder“ angeboten, das in Ergänzung der Daten zum aktuellen Spielgeschehen beispielsweise historische Kontextdaten zu vergangenen Spielen liefert, auf die dann im Livekommentar eingegangen werden kann (DFL 2021).

Das berechenbare Spiel?

Die weitreichenden Möglichkeiten der KI-basierten Datenanalyse in Bezug auf das Fußballspiel werfen die Frage auf, ob sich das Fußballspiel in Richtung „Rasenschach“ (Schramm 2023) entwickeln könnte, also Spielverläufe mehr oder weniger vorhersagbar werden. Bisher macht den Charme von Fußball wesentlich aus, dass er mit seinen wenigen Toren ein stark vom Zufall geprägtes Spiel ist. Rund 50 % der Tore gehen auf zufällige Ereignisse zurück, etwa Schiedsrichter(Fehl)entscheidungen, Pfosten- und Lattentreffer oder ausrutschende Abwehrspieler/innen, die nicht in direktem Zusammenhang mit dem Können einzelner Spieler/innen oder des gesamten Teams stehen (Wunderlich 2022, S.281 f.). Zwar lässt sich auf Basis von Daten aus der englischen Premier League aus den Spielsaisons 2012/2013 bis

2018/2019 ableiten, dass über die 7 Jahre hinweg der Zufall bei der Torerzielung abgenommen hat, jedoch ist er nach wie vor kennzeichnend für das Fußballspiel (Wunderlich et al. 2021). Nichtsdestotrotz können die neuen KI-basierten Analyse- und Interventionswerkzeuge dazu beitragen, möglichst wenig im Spielgeschehen dem Zufall zu überlassen. Wenngleich es einige Ansätze zur Vermessung von Spieler/innen und Mannschaften sowie Taktiken gibt, sind manche Fragen noch gänzlich unbeantwortet, etwa die nach den Qualitäten eines/einer Manager/in und deren Auswirkungen auf den Teamerfolg (Smith 2023).

Mögliche vertiefte Bearbeitung des Themas

Das Themenkurzprofil bietet einen Überblick über die zentralen Entwicklungen im Bereich Big Data und KI in Bezug auf das Fußballspiel. Vor dem Hintergrund dieser Entwicklungen bestehen zurzeit keine politischen Handlungsbedarfe, sodass eine vertiefende Analyse des Themas nicht angezeigt ist. Durch eine voraussichtlich stärkere Verbreitung der Datenerfassung im Jugend- und Amateurbereich könnte sich dies jedoch ändern, sodass das Thema zu einem späteren Zeitpunkt mit einer stärkeren Fokussierung auf Aspekte wie Datenschutz und digitale Souveränität weiter bearbeitet werden könnte.

Literatur

- ▶ Anzer, G.; Bauer, P.; Höner, O. (2022): Maschinelles Lernen mit Spieldaten: Die Erkennung von Gegenpressing im Fußball. In: Memmert, D. (Hg.): Spielanalyse im Sportspiel. Berlin/Heidelberg, S.255–264
- ▶ Biermann, C. (2018): Matchplan. Die neue Fußball-Matrix. Köln
- ▶ Brüggemann, K. (2022): Die Fußballblase. Hinter den Kulissen eines Milliardengeschäfts. Berlin/Heidelberg
- ▶ Brunnschneider, M.; Hahn, M. (2022): Scouting. In: Memmert, D. (Hg.): Spielanalyse im Sportspiel. Berlin/Heidelberg, S.183–191
- ▶ Bundesliga (2020a): DFL und Amazon Web Services stellen neue Echtzeit-Spielanalysen bereit. <https://www.bundesliga.com/de/bundesliga/news/neue-echtzeit-spielanalysen-der-dfl-und-amazon-web-services-11247> (11.6.2024)
- ▶ Bundesliga (2020b): Realformation als neue Echtzeit-Statistik der „Bundesliga Match Facts powered by AWS“. <https://www.bundesliga.com/de/bundesliga/news/realformation-position-aufstellung-taktik-aws-match-facts-echtzeit-statistik-daten-11481> (11.6.2024)
- ▶ Bundesliga (2021b): Was sind eigentlich Expected Goals? Eine Erklärung des xGoals-Modells. <https://www.bundesliga.com/de/bundesliga/news/expected-goals-xgoals-als-fussball-analyse-statistik-3760> (11.6.2024)
- ▶ Bundesliga (2022): Neuer Bundesliga Match Fact: Leistung unter Gegnerdruck. <https://www.bundesliga.com/de/bundesliga/news/match-facts-powered-by-aws-leistung-unter-gegnerdruck-amazon-analyse-stats-21389> (11.6.2024)
- ▶ Bundesliga (2023a): Bundesliga Match Facts powered by AWS: Shot Speed ist Match Fact Nummer 15. <https://www.bundesliga.com/de/bundesliga/news/bundesliga-match-facts-aws-schussgeschwindigkeit-shot-speed-25238> (11.6.2024)
- ▶ Bundesliga (2023b): Torwart-Effizienz und Balleroberungsdauer als neue Bundesliga Match Facts. <https://www.bundesliga.com/de/bundesliga/news/match-facts-torwart-effizienz-ball-eroberungsdauer-analyse-aws-amazon-dfl-23069> (11.6.2024)
- ▶ Bundesliga (o.J.a): Bundesliga Statistiken 2023–2024. <https://www.bundesliga.com/de/bundesliga/statistiken/bmf-zone> (11.6.2024)
- ▶ Bundesliga (o.J.b): Hol dir die App! <https://www.bundesliga.com/de/bundesliga/app> (11.6.2024)
- ▶ Clemente, F.; Martins, F.; Kalamaras, D.; Del Wong, P.; Mendes, R. (2015): General network analysis of national soccer teams in FIFA World Cup 2014. In: International Journal of Performance Analysis in Sport 15(1), S.80–96
- ▶ DFB (Deutscher Fußball-Bund) (2023): DFB und Sportec Solutions weiten Zusammenarbeit aus. <https://www.dfb.de/3-liga/news-detail/dfb-und-sportec-solutions-weiten-zusammenarbeit-aus-254872/> (11.6.2024)
- ▶ DFL (2021): Data Story Finder – Künstliche Intelligenz unterstützt den Live-Kommentar, <https://www.dfl.de/de/innovation/data-story-finder-kuenstliche-intelligenz-unterstuetzt-den-live-kommentar/> (11.6.2024)
- ▶ DFL (o.J.): Der Ursprung innovativer Statistiken: Die Offiziellen Spieldaten. <https://www.dfl.de/de/hintergrund/spieldaten/der-ursprung-innovativer-statistiken-die-offiziellen-spieldaten/> (11.6.2024)
- ▶ Diez, E. (2022): Was Daten mit dem Fußball machen. <https://www.swr.de/wissen/was-daten-mit-dem-fussball-machen-108.html> (11.6.2024)
- ▶ Engelen, M. (2024): Mehr Tore dank KI? Google Deepmind und der FC Liverpool entwickeln TacticAI. <https://t3n.de/news/google-deepmind-tacticalai-liverpool-tore-ki-eckballe-1615090/> (11.6.2024)
- ▶ FIFPRO (Fédération Internationale des Associations de Footballeurs Professionnels) (2022): Charter of Player Data Rights launched for professional footballers. <https://fifpro.org/en/supporting-players/competitions-innovation-and-growth/player-performance-data/charter-of-player-data-rights-launched-for-professional-footballers> (11.6.2024)
- ▶ FIFPRO (2023): FIFPRO, unions commit to create centralised player data management platform. <https://fifpro.org/en/supporting-players/competitions-innovation-and-growth/player-performance-data/fifpro-member-unions-commit-to-create-centralized-player-data-management-platform-for-the-football-industry> (11.6.2024)
- ▶ Garnica Caparrós, M. (2022): Key Performance Indicators – KPIs. In: Memmert, D. (Hg.): Spielanalyse im Sportspiel. Berlin/Heidelberg, S.173–180

- ▶ Hartley, N. (2021): Professional footballers threaten data firms with GDPR legal action. <https://www.bbc.com/news/uk-wales-58873132> (11.6.2024)
- ▶ Hofmann, B. (2024): Datenexperte Graham im Interview: „Salah wäre nicht Jürgens erste Wahl gewesen“. <https://www.kicker.de/datenexperte-graham-im-interview-salah-waere-nicht-juergens-erste-wahl-gewesen-1001869/artikel> (11.6.2024)
- ▶ Huber, O. (2023): Wie Algorithmen den Fußball verändern. Ein Spiel der Daten. <https://www.ardalpha.de/wissen/fussball-digital-analyse-algorithmus-statistik-zahlen-spiel-100.html> (11.6.2024)
- ▶ Hummel, T.; Karon, J. (o.J.): KI im Fußball – Mit Algorithmen zum Sieg. 21.7.2021, <https://www.swr.de/swrkultur/wissen/ki-im-fussball-mit-algorithmen-zum-sieg-104.html> (11.6.2024)
- ▶ Klingelhöfer, C. (2023): Trainingsrevolution: So verändert KI den Fußball. <https://www.ispo.com/sportstech/ki-im-sport-kuenstliche-intelligenz-im-fussball> (11.6.2024)
- ▶ Lutz, J.; Memmert, D.; Raabe, D.; Dornberger, R.; Donath, L. (2020): Wearables for Integrative Performance and Tactic Analyses: Opportunities, Challenges, and Future Directions. In: *International journal of environmental research and public health* 17(1), Art. 59
- ▶ Memmert, D. (2022): Spielanalyse im Jahr 2020. In: Memmert, D. (Hg.): *Spielanalyse im Sportspiel*. Berlin/Heidelberg, S.3–12
- ▶ Raabe, D. (2022): KPIs in der Bundesliga. In: Memmert, D. (Hg.): *Spielanalyse im Sportspiel*. Berlin/Heidelberg, S.237–243
- ▶ Schramm, M. (2023): KI und Sport - Wenn Fußball gläsern wird. <https://www.ardaudiothek.de/episode/iq-wissenschaft-und-forschung/ki-und-sport-wenn-fussball-glaesern-wird/bayern-2/94622850/> (11.6.2024)
- ▶ Seoudy, S. (2023): The Fight for Control of Performance Data: Is Project Red Card Headed for an Early Bath? <https://www.armstrongteasdale.com/thought-leadership/the-fight-for-control-of-performance-data-is-project-red-card-headed-for-an-early-bath/> (11.6.2024)
- ▶ Smith, R. (2023): Can this man make soccer smarter? <https://www.nytimes.com/2023/10/14/world/europe/liverpool-ian-graham-data.html> (11.6.2024)
- ▶ Transfermarkt.de (o.J.a): #4 Virgil van Dijk. 19.12.2023, <https://www.transfermarkt.de/virgil-van-dijk/profil/spieler/139208> (11.6.2024)
- ▶ Transfermarkt.de (o.J.b): #9 Philippe Coutinho. 2.5.2024, <https://www.transfermarkt.de/philippe-coutinho/profil/spieler/80444> (11.6.2024)
- ▶ Transfermarkt.de (o.J.c): #11 Mohamed Salah. 19.12.2023, <https://www.transfermarkt.de/mohamed-salah/profil/spieler/148455> (11.6.2024)
- ▶ VDV (Vereinigung der Vertragsfußballspieler e.V.) (o.J.): Die VDV – deine Spielergewerkschaft: Forderungen. <https://www.spielergewerkschaft.de/de/vdv/deine-spielergewerkschaft.htm> (11.6.2024)
- ▶ WDR 5 (2023): Ein anderes Spiel – Fußball und Digitalisierung. <https://www.ardaudiothek.de/episode/wdr-5-sport-inside-der-podcast-kritisch-konstruktiv-inklusiv/ein-anderes-spiel-fussball-und-digitalisierung/wdr-5/12391447/> (11.6.2024)
- ▶ Wunderlich, F. (2022): Grenzen der Spielanalyse. In: Memmert, D. (Hg.): *Spielanalyse im Sportspiel*. Berlin/Heidelberg, S.277–286
- ▶ Wunderlich, F.; Seck, A.; Memmert, D. (2021): The influence of randomness on goals in football decreases over time. An empirical analysis of randomness involved in goal scoring in the English Premier League. In: *Journal of sports sciences* 39(20), S.2322–2337

Das Horizon-Scanning ist Teil des methodischen Spektrums der Technikfolgenabschätzung im TAB.

**Horizon
SCANNING**

Mittels Horizon-Scanning werden neue technologische Entwicklungen beobachtet und diese systematisch auf ihre Chancen und Risiken bewertet. So werden technologische, ökonomische, ökologische, soziale und politische Veränderungspotenziale möglichst früh erfasst und beschrieben. Ziel des Horizon-Scannings ist es, einen Beitrag zur forschungs- und innovationspolitischen Orientierung und Meinungsbildung des Ausschusses für Bildung, Forschung und Technikfolgenabschätzung zu leisten.

In der praktischen Umsetzung werden im Horizon-Scanning softwaregestützte Such- und Analyseschritte mit expertenbasierten Validierungs- und Bewertungsprozessen kombiniert.

Herausgeber: Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB)

Gestaltung und Redaktion: VDI/VDE Innovation + Technik GmbH

Bildnachweise: Who_I_am/iStock (S. 1); Hispanolistic/iStock (S. 2); Creativimages/iStock (S. 4)

ISSN-Internet: 2629-2874