

# Internationales Fußball Institut

Hochschulzertifikat Data Analytics im professionellen Fußball

## DATA BASED GOALPLAYER SCOUTING

Abschlussarbeit

vorgelegt von  
Janik Backhaus

Dietrich Bonhoeffer Straße 29  
83043 Bad Aibling  
Tel.: 01734010191

Programmleiter: Dr. Malte Siegle  
Tag der Einreichung: 08.04.2023

# Inhaltsverzeichnis

- 1 Einleitung ..... 4**
- 2 Theoretischer Hintergrund: Datenanalyse..... 5**
  - 2.1 Entstehung und Ursprung Datenanalyse .....5
  - 2.2 Aufgabengebiete eines Data Analysts .....9
  - 2.3 Abgrenzung Data Analytics vs. Data Science ..... 11
- 3 Einordnung in Trainingswissenschaft ..... 12**
  - 3.1 Anwendungsgebiete für Data Analytics und Stand der Wissenschaft ..... 12
  - 3.2 Datenarten und deren Informationsgehalt..... 17
  - 3.3 Datenerhebungsmethoden, deren Eigenschaften und Qualitätsmerkmale ..... 19
- 4 Anforderungsprofil des Torspielers ..... 23**
  - 4.1 Vom Torspieler 1.0 zum Torspieler 4.0 .....23
  - 4.2 Torspieler spezifische Daten und Advanced Metrics.....27
- 5 Methodik..... 31**
  - 5.1 Wahl des Anwendungsgebietes & konkrete Zielstellung des datenanalytischen Ansatzes .....32
  - 5.2 Beschreibung der Datensätze.....33
  - 5.3 Berechnung von KPI´s und Clusteranalyse zur Kategorisierung von Torspielern .....35
- 6 Ergebnisse und Diskussion ..... 37**
  - 6.1 Darstellung mit Hilfe von Tableau und Erläuterung der gewonnenen Ergebnisse .....37
    - 6.1.1 Clusteranalyse nach Leistungsfähigkeit.....37
    - 6.1.2 Clusteranalyse nach Torspielertyp.....40
    - 6.1.3 Detailanalyse Zielverteidigung .....41
    - 6.1.4 Detailanalyse Mitspielverhalten bei gegnerischem Ballbesitz.....49
    - 6.1.5 Detailanalyse Mitspielverhalten im eigenen Ballbesitz .....54
  - 6.2 Informationsgehalt der Ergebnisse .....62
- 7 Ausblick ..... 66**

7.1	Mögliche Erweiterung des eigenen Ansatzes.....	66
7.1.1	Optimierungspotenziale Spieldatenerfassung - Zielverteidigung.....	67
7.1.2	Optimierungspotenziale Spieldatenerfassung – Mitspielverhalten im eigenen Ballbesitz .....	69
7.1.3	Optimierungspotenziale Spieldatenerfassung – Mitspielverhalten bei gegnerischem Ballbesitz .....	71
7.2	“Wo geht die Reise hin im Datenanalysebereich”?.....	73
<b>8</b>	<b>Literaturverzeichnis .....</b>	<b>75</b>
<b>9</b>	<b>Abbildungsverzeichnis .....</b>	<b>78</b>

## 1 Einleitung

Im professionellen Fußball gewinnen Daten und Datenanalyse zunehmend an Bedeutung, um fundierte Entscheidungen zu treffen und langfristig erfolgreich zu sein. Während insbesondere das Work-Load-Management, zum Teil die Spielanalyse und das Scouting von Feldspielern bereits weit verbreitete Anwendungsbereiche für Daten im Fußball darstellen, ist das Scouting für die Position des Torspielers bislang noch stark von subjektiven Einschätzungen, überbewerteten physischen Aspekten wie der Körpergröße und unzureichender Kontextualisierung von Daten geprägt.

Die mediale Berichterstattung fokussiert sich oft nur auf eine begrenzte Anzahl an Leistungsindikatoren aus dem Teilbereich der Zielverteidigung wie Paraden, Fangquote oder Clean Sheets, ohne den Kontext der Torspieler-Aktionen im Spiel angemessen zu berücksichtigen (z.B. Qualität des Schusses). 94% aller Torspieler-Aktionen pro Spiel, die sich auf das Mitspielverhalten im eigenen Ballbesitz und das Mitspielverhalten bei gegnerischem Ballbesitz (Raumverteidigung) verteilen, bleiben in Gänze unberücksichtigt. Obwohl führende Anbieter in ihren Matchreports bereits grafische Darstellungen und Analysen für Torspielerleistungen anbieten, sind diese in ihrer Aussagekraft zum gegenwärtigen Zeitpunkt sehr begrenzt. In Vereinen und Verbänden fehlt es zudem häufig an Experten, die sowohl über ausreichendes Detailwissen über die Position verfügen als auch die Daten korrekt zu interpretieren können (Datenanalysten mit ausreichender fachlicher Expertise oder Torspieler-Trainer/„Legenden“/Scouts mit Affinität und ausreichendem Verständnis von Daten). Der zielgerichtete Einsatz von Daten im Torspieler-Scouting bietet jedoch zweifelsohne die Möglichkeit, den Prozess effizienter und effektiver zu gestalten um potenzielle Torspieler, die in das jeweilige Mannschaftskonzept passen, schneller zu identifizieren.

In dieser Arbeit sollen deshalb Lösungsansätze für ein datenbasiertes Torspieler-Scouting vorgestellt werden, die das technisch-taktische Anforderungsprofil des Torspielers 4.0 in den Kernbereichen Gegentore verhindern (Zielverteidigung), Mitspielverhalten im eigenen Ballbesitz und Mitspielverhalten bei gegnerischem Ballbesitz (Raumverteidigung) erstmals ganzheitlich betrachten.

Um diesen innovativen Ansatz zu entwickeln und den Prozess für den Leser besser zugänglich zu machen, werden in dieser Arbeit zunächst verschiedene theoretische Grundlagen erläutert. Kapitel 2 beschäftigt sich mit dem theoretischen Hintergrund der Datenanalyse, einschließlich der Entstehung und Ursprünge der Datenanalyse, den Aufgabenbereichen eines Data Analysts sowie

einer Abgrenzung zwischen Data Analytics und Data Science. Kapitel 3 beschreibt die Anwendungsgebiete von Datenanalyse im Fußball im Kontext der Trainingswissenschaft und gibt einen Überblick über Datenarten, Datenerhebungsmethoden und deren Qualitätsmerkmale.

Kapitel 4, 5 und 6 bilden den Hauptteil der Arbeit. In Kapitel 4 wird das Anforderungsprofil des Torspielers analysiert und seine historische Entwicklung bis zum Torspieler 4.0 beschrieben. Zudem werden torspielerspezifische Daten und Advanced Metrics vorgestellt und kategorisiert. Kapitel 5 beschreibt die Methodik des datenbasierten Torspieler-Scoutings, erläutert die konkrete Zielstellung, erklärt die verwendeten Datensätze und stellt die Berechnung von torspielerspezifischen Key Performance Indicator(s) (KPI)'s und die Clusteranalyse zur Kategorisierung von Torspielern als zentrale Elemente des Prozesses vor. In Kapitel 6 werden die Ergebnisse der datenbasierten Torspieleranalyse mit Hilfe von Tableau dargestellt, erläutert und interpretiert sowie deren Informationsgehalt diskutiert. Mögliche Erweiterungen des Ansatzes und eine Einschätzung zur zukünftigen Entwicklung im Bereich der Datenanalyse bilden den Ausblick der Arbeit.

## **2 Theoretischer Hintergrund: Datenanalyse**

Das Kapitel beginnt mit einem Überblick über die Entstehung und den Ursprung der Datenanalyse und beschreibt wie sich diese Disziplin im Laufe der Zeit entwickelt hat. Anschließend werden die Aufgabenbereiche eines Data Analysts beschrieben, um zu verstehen, welche Fähigkeiten und Kenntnisse erforderlich sind, um Daten erfolgreich zu analysieren und zu interpretieren. Darüber hinaus wird zum besseren Verständnis eine Abgrenzung zwischen Data Analytics und Data Science vorgenommen.

### **2.1 Entstehung und Ursprung Datenanalyse**

Die Geschichte der Datenanalyse im Fußball reicht bis zu den Anfängen des Sports zurück, als Trainer und interessierte Beobachter darüber nachdachten, wie man erfolgreich spielen kann. In den frühen Jahren bediente man sich mangels technischer Hilfsmittel noch altbewährter Werkzeuge wie Zettel und Stift, um Leistungen von Spielern zu dokumentieren. „Aus zunächst undurchsichtigen Kritzeleien zu Spielzügen und Passstafetten entstanden über die Jahre sogenannte Handnotationssysteme“ (Memmert & Raabe, 2017, S. 24) – rückblickend die Geburtsstunde der heutigen Spielanalyse.

Der Beginn der Handnotationssysteme liegt allerdings nicht im Bereich des Fußballs, sondern im Baseball. Die amerikanischen Sportarten, einschließlich Baseball, Football, Eishockey und Basketball, bieten aufgrund ihrer vergleichsweise häufigen und standardisierten

Unterbrechungen sowie ihrer weniger komplexen Spielstruktur einfachere Möglichkeiten zur Erfassung und Quantifizierung von Spielaktionen. Die fließenden und dynamischen Übergänge der Spielphasen und die nichtlineare Natur des Spielgeschehens im Eishockey und Basketball erschweren zweifelsohne die Abgrenzung der einzelnen Aktionen. Aufgrund der Häufigkeit der standardisierten Unterbrechungen, der geringeren Anzahl der Spieler und nicht zuletzt der deutlich höheren Anzahl der potenziell spielentscheidenden Aktionen sind sie dennoch weniger anspruchsvoll als die Spiel- und Datenanalyse im Fußball. Aus diesem Grund nehmen diese Sportarten eine Vorreiterrolle in der Geschichte der Datenanalyse im Sport ein. Zusammenfassend kann festgestellt werden, dass zeitverzögert immer wieder versucht wird die genutzten Prozesse, Technologien und Erkenntnisse für den Kontext komplexerer Sportspiele wie Fußball zu adaptieren. Im Folgenden sollen die wichtigsten Meilensteine der historischen Entwicklung kurz erläutert werden, da diese auch im Fußball von herausragender Bedeutung sind.

- Der Amerikaner Hugh Fullerton veröffentlichte bereits 1912 einen Aufsatz, in dem er das Baseballfeld in mehrere Zonen einteilte und über 10.000 in die Zonen geschlagene Bälle auf Erfolgswahrscheinlichkeiten analysierte (vgl. Memmert & Raabe, 2017, S. 24).
- Lloyd Messersmith entwickelte in den 1930er Jahren eines der ersten Handnotationssysteme im Sport (Basketball, Baseball, Football). Er dokumentierte nicht nur Spielaktionen, sondern auch die physischen Leistungen (u.a. Laufdistanz) der Athleten. Die Messung erfolgte mit einem selbst entwickelten elektrischen "Verfolgungsapparat", der aus einem Miniaturspielfeld, einer Nadel zur Aufzeichnung des Laufweges, einem elektrischen Impulszähler sowie einem Batteriekasten bestand (vgl. ebd., S. 25ff.).
- Der Statistiker Bill James hinterfragte in den 1970er Jahren traditionelle Statistiken im Baseball und entwickelte neue Analysen, die unter dem Namen Sabermetrics bekannt wurden. Ziel war es die Spieler objektiver zu bewerten und die Gesamtleistung eines Spielers zu berücksichtigen, anstatt nur Teilaspekte zu betrachten. Die Ideen von Sabermetrics wurden weiterentwickelt, erlebten aber erst 2002 ihren großen Durchbruch, als der Manager des Proficlubs Oakland Athletics, Billy Bean, Sabermetrics-Maßzahlen nutzte, um seinen Kader zusammen zu stellen und damit großen Erfolg hatte. Diese Erfolgsgeschichte ist heute vor allem durch den Film „Moneyball“ bekannt. Chen et. al. (2010) beschreiben Sabermetrics als Wandel von traditionellen Statistiken hin zum Wissensmanagement (vgl. Chen et. al., 2010, S. 36). Gavin Fleig, Leiter der Analyseabteilung von Manchester City würdigte James´ Verdienste in dem er sinngemäß betont, dass es einen Bill James des Fußballs brauche um die Spiel-/Datenanalyse zu revolutionieren (vgl. Memmert & Raabe, 2017, S. 88ff.).

- In den 1980er Jahren startete der Statistiker Dean Oliver eine ähnliche Datenrevolution im Basketball (vgl. Chen et. al., 2010, S. 36)

Zusammenfassend bedeutet das Sabermetrics-Konzept in seiner ursprünglichsten Form die systematische Analyse von Daten zur Identifikation von Spielern oder Talenten, die in der Praxis unterbewertet sind oder ihr Potenzial im aktuellen Verein noch nicht voll ausschöpfen. Der Transfer dieser Spieler zu vergleichsweise günstigen Konditionen dient dem primären Ziel, selbst erfolgreich zu sein, während auch die Möglichkeit des späteren Verkaufs mit einem signifikanten Transfergewinn eine entscheidende Rolle spielt.

Die Implementierung eines datenbasierten Spielerscoutings, welches auf den Konzepten der Sabermetrics beruht, wurde im Bereich des Fußballs erstmals durch den Multiclub-Owner Matthew Benham in seinen Vereinen FC Midtjylland und FC Brentford Mitte der 2010er-Jahre eingeführt (vgl. transfermarkt.de, 2021). Backhaus (2022) stellt in diesem Zusammenhang fest, dass sich Multiclub-Ownerships, bei denen es durch den Austausch von Informationen und Wissen zur vollständigen Entfaltung positiver Clustereffekte kommt, einen Wettbewerbsvorteil erarbeitet haben (vgl. Backhaus, 2022, S.85; S.92). Ein weiteres Beispiel hierfür ist die von Investoren aus den USA und China finanzierte Pacific Media Group, bei der Billy Bean stiller Teilhaber ist, die vor allem durch den FC Barnsley und die Meisterschaft des OGC Nizza bekannt geworden ist. Durch den Einsatz von komplexen Datenanalyse-Methoden identifiziert die Investorengruppe unterschätzte Spieler und hat auf diese Weise nahezu in allen Vereinen, an denen sie beteiligt ist (darunter AS Nancy, KV Ostende, FC Den Bosch, Esbjerg fB, FC Thun und 1. FC Kaiserslautern), zumindest kurzfristigen, sportlichen Erfolg, während sich zugleich die Marktwerte der Einzelspieler signifikant erhöhten (vgl. Slater, 2021). Die Aufzählung von Vereinen, deren jüngste Erfolge ebenfalls primär auf datenbasiertes Scouting zurückzuführen sind, könnte durch weitere namhafte Vereine wie den FC Liverpool, FC Arsenal, Brighton, AC Milan, AZ Alkmaar, Union SG sowie den FC Toulouse, um nur einige der schnell wachsenden Liste zu nennen, erweitert werden. Allerdings würde eine detaillierte Darstellung der Ansätze, die sich in Herangehensweise nur minimal unterscheiden, den Rahmen dieser Arbeit sprengen.

In den Anfängen des Sabermetrics-Konzepts im Fußball wurde hauptsächlich der Vergleich von erwarteten Toren (xGoals) und tatsächlich erzielten Toren genutzt, um unterbewertete Spieler zu identifizieren. Im Verlauf der Zeit hat sich die Methodik dahingehend weiterentwickelt, dass Daten nun dazu beitragen, Spieler zu erkennen, die aufgrund ihrer Fähigkeiten in das Anforderungsprofil eines spezifischen Spielsystem passen, sowie vorherzusagen, wie diese Spieler miteinander interagieren werden. Gegenwärtig werden von den meisten Profivereinen hybride Scoutingstrategien verwendet, bei denen eine initiale Vorauswahl oder die Selektion angebotener

Spieler anhand quantitativer Daten und Metriken erfolgt, die anschließend mit qualitativen Methoden verfeinert wird.

Die Entwicklung hin zu diesen Strategien war jedoch langwierig. Ursprünglich war die Leistungsbeurteilung im Fußball, insbesondere im Kontext der Spielanalyse und des Scoutings, hauptsächlich von qualitativen "traditionellen" Spielbeobachtungsverfahren geprägt. Seit den 2010er-Jahren haben jedoch verstärkt Daten Einzug in den Fußball erhalten, um diese Verfahren zu unterstützen und Beobachtungs-/Beurteilungsverzerrungen, sogenannte Bias, zu minimieren. Memmert und Raabe (2017) gliedern die historische Entwicklung in vier Phasen, die in Abbildung 1 dargestellt sind.

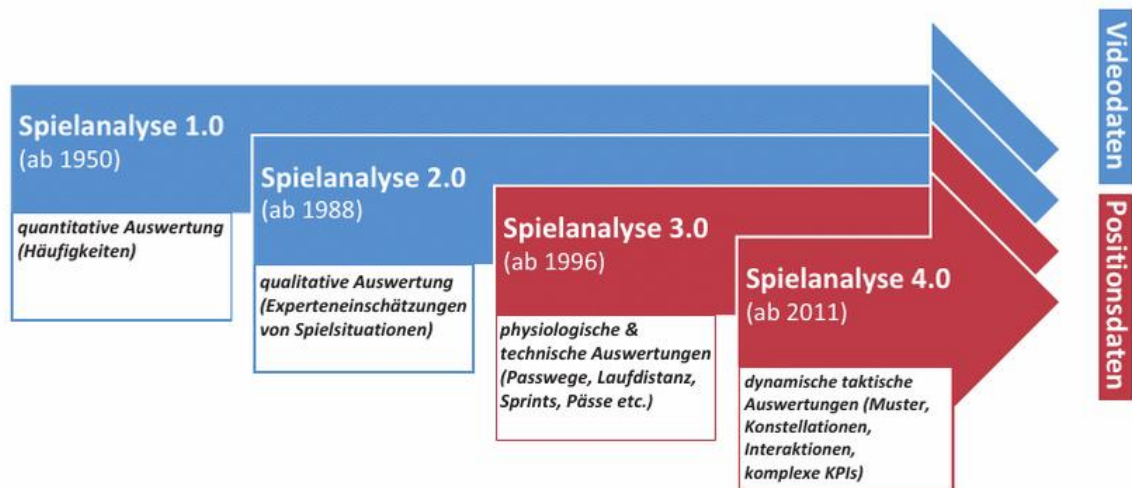


Abbildung 1. Von der Spielanalyse 1.0 zur Spielanalyse 4.0. Quelle: Memmert & Raabe, 2017, S.8

Ausgehend von der Spielanalyse 1.0, die sich auf die bereits beschriebene quantitative Auswertung von Häufigkeiten durch einfache Handnotationssysteme konzentrierte, änderten sich im Laufe der Zeit die Art, Erhebung und Nutzung von Daten bis hin zur Entwicklung komplexer Leistungsindikatoren (Spielanalyse 4.0).

Der Brite Charles Reep kann als Pionier der fußballspezifischen Datenanalyse bezeichnet werden. Im Rahmen seiner Forschung analysierte er knapp 2500 Spiele mit Hilfe eines eigens entwickelten Notationssystems. Bereits im Jahr 1968 gelangte er zu dem Schluss, dass 80% der Tore nach Ballbesitzphasen von weniger als 3 Pässen erzielt werden. Dieses Erkenntnis hat bis heute einen maßgeblichen Einfluss auf den typisch englischen Kick and Rush-Stil (vgl. Memmert & Raabe, 2017, S. 28ff). Frühe Spielanalysten begannen in der Folge, das Spielfeld ähnlich wie im Baseball in Zonen zu unterteilen und feine Rasterfelder zu entwickeln, um Ballaktionen



präziser zu lokalisieren. Diese Rasterfeldanalysen stellen den historischen Vorläufer der heutigen Heatmaps dar, die zur Visualisierung der Positionen der Spieler auf dem Spielfeld eingesetzt werden. Der Schotte A.H. Ali entwickelte im Jahr 1986 auf dieser Basis ein Handnotationssystem, mit dem er Angriffsaktionen mittels X-Y-Koordinaten lokalisierte und somit einen ersten Vorgänger der heutigen Positionsdaten-Analyse erschuf (vgl. ebd., S.31f). In den 1990er Jahren wurde die Spielanalyse dann fast vollständig digitalisiert, wobei vor allem die Dateneingabe beschleunigt wurde, indem Tastaturen und Spracherkennungssysteme eingesetzt wurden. Mit Hilfe von stationären Kameras wurde das Spielfeld aufgenommen, notierte Informationen zeitlich codiert und mit gespeicherten Matchvideos synchronisiert – praktisch der Beginn des heute bekannten Taggings (vgl. ebd., S.33ff).

Die Erhebung von Positionsdaten wurde erst durch den technologischen Fortschritt möglich. 1984 gewannen Wissenschaftler aus Flandern aus einem Spiel erstmals einen Positionsdatensatz. Die kommerzielle Vermarktung von Tracking-Technologien begann gut zehn Jahre später (vgl. ebd., S. 44ff). Aktuell werden in der Bundesliga allein durch die DFL-Tochter-Firma Sports-Tec-Solutions pro Spiel ca. 3,6 Millionen Positionsdatenpunkte und 1.600 Spielereignisse erfasst (vgl. DFL, 2020). Der Markt der Tracking-Systemen-Anbieter ist inzwischen hart umkämpft und unterliegt einer dynamischen Veränderung. Auf die verschiedenen Technologien zur Erfassung von Positionsdaten, zu deren Entwicklung auch die Militärforschung erheblich beigetragen hat beigetragen, wird in 3.3. genauer eingegangen.

Zusammenfassend sind Daten in der heutigen Zeit im Profifußball unverzichtbar und werden in verschiedenen Bereichen wie der taktischen Analyse des kommenden Gegners, der Verbesserung der eigenen Spielweise, im Scouting oder zur optimalen Belastungssteuerung eingesetzt.

## **2.2 Aufgabengebiete eines Data Analysts**

Da das Berufsfeld des Data Analysts im Profifußball noch vergleichsweise jung ist, Strukturen und Budgets der Vereine sowie die Kompetenzen und Spezialgebiete der Personen, die diese Rolle übernehmen, stark variieren, ist es schwierig, eine allgemeingültige Definition der Aufgabengebiete zu geben.

Fest steht, dass Datenanalysten aufgrund ihrer Kernaufgaben in der Analyse und Interpretation von Spiel- und Spielerdaten sowie der Generierung objektiver Informationen zu Leistungsentwicklungen und -potenzialen nahezu unverzichtbar sind. Zu den Kernaufgaben des Data Analysts zählt darüber hinaus das kurz-, mittel- und langfristige Datenmanagement (Dokumentation, Speicherung, Verarbeitung) sowie die leicht verständliche Aufbereitung seiner Ergebnisse in Form von Reportings oder Präsentationen, die in der Kommunikation mit den

verschiedenen Zielgruppen (Trainer, Sportdirektor, Vorstand...) durch Visualisierungen unterstützt werden sollten. Abhängig von der Vereinsstruktur können mit der Datenerfassung, einschließlich operativer Abläufe im Tagesgeschäft und der Auswahl geeigneter Anbieter, Elemente aus dem Data Engineering zum Aufgabenbereich des Datenanalysten gehören. Komplexe Aufgaben wie beispielsweise datenbasierte Vorhersagen, die üblicherweise einem Data Scientist zugeordnet werden, können ebenfalls Teile des Aufgabenbereichs eines Data Analysts sein, sofern keine eigenen Data Scientists im Verein beschäftigt werden und komplexe Fragestellungen kostengünstig intern gelöst werden sollen.

Im sportlichen Bereich existieren drei Hauptaufgabenbereiche, welche vom Data Analyst übernommen werden müssen, sofern keine weiteren Datenanalysten oder spezialisierten Bereichsexperten (Trainer, Physiotherapeuten, Spielanalysten, Scouts) mit ausreichender Datenanalysekompetenz vorhanden sind: Spielanalyse, Workload-Management und Scouting. Ein vierter Aufgabenbereich, die Vor-/Nachbereitung und die Begleitung des Trainings, ist ebenfalls relevant, jedoch gibt es große Überlappungen mit dem Workload-Management und einigen Teilen der Spielanalyse im Trainingskontext, weshalb dieser hier nicht weiter vertieft wird.

Im Bereich der Spielanalyse werden im operativen Geschäft vier verschiedene Arten der Datenanalyse unterschieden: die Pre-Match-Analyse, die Live-Match-Analyse, die Post-Match-Analyse sowie die langfristige Analyse der Entwicklung der eigenen Mannschaft. Diese Kategorien lassen sich in weitere Teilbereiche wie das physische Leistungsprofil (Athletik), die Systeme (Mannschaftstaktik), die Spielidee (Gruppentaktik) und die Spieler (Individualtaktik) differenzieren. Der Datenanalyst hat hierbei die Aufgabe, die umfangreichen Datenmengen sinnvoll zusammenzufassen. Dadurch kann eine höhere Detailtiefe und objektivere Informationen im Vergleich zur subjektiven Wahrnehmung des Spielanalysten erreicht werden. Dies ermöglicht es, Entwicklungen im zeitlichen Verlauf des Spiels einfacher zu erkennen und Vorhersagen oder taktische Lösungsmöglichkeiten für den weiteren Spielverlauf zu entwickeln.

Im Bereich des Workload-Managements geht es primär darum, die Belastung der Spieler datenbasiert zu steuern, um sowohl eine optimale Verletzungsprävention zu gewährleisten als auch die bestmögliche Matchfitness zu erreichen.

Last but not least beinhaltet das Aufgabenprofil des Data Analysts das Scouting. Traditionelles Live-Scouting erfordert viel Zeit und Ressourcen, und Einschätzungen können trotz erfahrener Scouts oft subjektiv sein und Wahrnehmungsverzerrungen enthalten. Daher gewinnen objektive Daten immer mehr an Bedeutung, um bei der Verpflichtung neuer Spieler eine Risikominimierung oder -abschätzung zu ermöglichen, eine weltweite Abdeckung von Ligen zu erreichen, Zeit zu sparen und im Livescouting zeiteffizienter zu arbeiten. Der Markt bietet eine Vielzahl von Anbietern, daher ist es eine zentrale Aufgabe des Datenanalysten, die Vor- und Nachteile der

unterschiedlichen Datenpakete/-anbieter sowie die damit verbundenen Kostenstrukturen im Detail zu kennen. Dies ermöglicht ihm bzw. seinem Team, genau diejenigen Daten zu erwerben, mit denen er die verschiedenen Positionsprofile und vereinspezifischen KPIs am besten analysieren kann, sofern er diese bereits abteilungsübergreifend auf Datenbasis definiert hat.

### **2.3 Abgrenzung Data Analytics vs. Data Science**

Um die Aufgabengebiete des Data Analysts im Detail zu verstehen ist es notwendig das im Fußballkontext häufig gemeinsam vorkommende Begriffspaar Data Analytics und Data Science genauer zu erläutern.

Es ist wichtig zu betonen, dass Data Science und Data Analytics nicht als eigenständige Bereiche zu verstehen sind. Vielmehr ist der Begriff "Datenanalyse" im Deutschen eine generalisierende Überkategorie für die allgemeine Untersuchung von Daten. Data Science ist somit ein Teilbereich von Data Analytics, was bedeutet, dass die Rollen des Data Scientist und des Data Analyst eng miteinander verwandt sind. Im Wesentlichen geht es für beide darum, Erkenntnisse wie Korrelationen, Kausalitäten und Muster aus Datensätzen, die zunächst bereinigt und in BI-Tools integriert werden müssen um valide damit arbeiten zu können, zu gewinnen und daraus Handlungsempfehlungen für die Praxis abzuleiten (vgl. Wunderlich, 2023). Was für beide ein grundlegendes Fußballverständnis voraussetzt um erkennbare Zusammenhänge auch zu erfassen und richtig einzuordnen.

Im Vergleich zu Data Science befasst sich Data Analytics, überwiegend retrospektiv (mit Daten aus der Vergangenheit), mit der Extrahierung von vergleichsweise übersichtlichen Datensätzen. Die Tätigkeiten eines Data Analysts sind daher stark anwendungsbezogen (vgl. 2.2) und erfordern neben fußballspezifischem Fachwissen sowie Kenntnissen in Sportwissenschaft und Statistik vor allem Fähigkeiten im Umgang mit verhältnismäßig übersichtlichen und gut strukturierten Datensätzen sowie deren Analyse und Visualisierung. Die wichtigste Fähigkeit eines Data Analysts besteht darin, auf Basis der Daten aussagekräftige Erkenntnisse zu gewinnen, die Entscheidungsträger wie Trainer oder Sportdirektoren nutzen können um Entscheidungen zu treffen, und diese in visuell ansprechenden und verständlichen Formaten zu präsentieren.

In Abgrenzung dazu beschäftigt sich Data Science mit den wissenschaftlichen Grundlagen der Mustererkennung und Klassifizierung, wobei Data Scientists mit Hilfe verschiedenster statistischer Methoden prädiktive Modelle entwickeln, um Vorhersagen und Empfehlungen für die Zukunft abzuleiten. Im Gegensatz zu Data Analytics ist Data Science sehr viel programmierintensiver und erfordert ein tieferes Verständnis für Datenverarbeitung in großem Stil. Data Scientists nutzen in der Regel weniger analytische Methoden, sondern komplexe

statistische Auswertungen sehr viel größer, oftmals unstrukturierter und ungeordneter, Datenmengen aus verschiedenen sportlichen und außersportlichen Bereichen, um diese mit Hilfe mathematischer Algorithmen zu modellieren und daraus Erkenntnisse zu gewinnen (vgl. Wunderlich, 2023).

Ein zentraler Aspekt beider Bereiche ist die Kommunikation innerhalb des Vereins. Data Analytics und Data Science agieren an der Schnittstelle zwischen Sportwissenschaft, Mathematik und Informatik. Die meisten Trainer, Sportdirektoren und Vorstände interessieren sich jedoch nicht für die genaue Herleitung statistischer Details, sondern wollen zuverlässige, leicht verständliche Ergebnisse und Erklärungen ihrer Bedeutung für sportlichen Entscheidungen. Die Arbeit des Data Scientists ist oft sehr komplex und enthält viele Fachbegriffe, die für Entscheidungsträger aus der Praxis schwer verständlich sind. Diese Verständnislücken, die andersrum bezüglich fußballerischen Fachwissens genauso bestehen, führen oft zu Barrieren zwischen den Entscheidungsträgern und Data Scientists. In diesem Kontext spielt der Data Analyst eine wichtige Rolle, da er in der Regel näher an der Praxis arbeitet und daher „die gleiche Sprache“ wie die Entscheidungsträger spricht. Gleichzeitig besitzt er das notwendige, grundlegende Know-how in Data Science, um die Thematik verständlich zu erklären und gegebenenfalls zu visualisieren.

### **3 Einordnung in Trainingswissenschaft**

Dieses Kapitel befasst sich mit der Einordnung von Data Analytics in die Trainingswissenschaft. Es werden verschiedene Anwendungsgebiete von Data Analytics im Kontext der trainingswissenschaftlichen Gegenstandsbereiche aufgezeigt und der aktuelle Stand der Wissenschaft präsentiert. Zudem werden die verschiedenen Datenarten und ihr Informationsgehalt sowie die Eigenschaften und Qualitätsmerkmale der Datenerhebungsmethoden erläutert.

#### **3.1 Anwendungsgebiete für Data Analytics und Stand der Wissenschaft**

Die Datenanalyse, welche als Teilbereich und Ausprägungsform der Spielanalyse zu sehen ist, ist der sportwissenschaftlichen Teildisziplin der Trainingswissenschaft zuzuordnen. Schnabel et. al (2016) definieren die Trainingswissenschaft als

„Disziplin der Sportwissenschaft, die das Beziehungsgefüge „sportliche Leistung/ Leistungsfähigkeit – sportliches Training“ unter den Zielaspekten der körperlichen Vervollkommnung, der motorischen bzw. der sportlichen Leistungsentwicklung sowie des Vergleichs im sportlichen Wettkampf zum Gegenstand hat. [...]“ (Schnabel et. al., 2016, S. 25)

Innerhalb der Trainingswissenschaft werden somit das sportliche Training, als langfristig geplanter und strukturierter Prozess von zentraler Bedeutung, sowie die sportliche Leistung(-sfähigkeit) als wesentlicher Ansatzpunkt des Trainings und der sportliche Wettkampf als Gegenstandsbereiche betrachtet (vgl. ebd., S.17f.). Bereits in dieser Formulierung wird offensichtlich, dass die gegenseitige Abhängigkeit und Wechselwirkung der einzelnen Gegenstandsbereiche von großer Bedeutung ist (vgl. Abb.2).



Abbildung 2: Wechselseitiger Zusammenhang der Gegenstandsbereiche der Trainingswissenschaft. Quelle: Schnabel et. al., 2016, S.19

Im Kontext von Sportspielen im Allgemeinen und Fußball im Besonderen ist es relevant, den interaktionalen Handlungscharakter zu berücksichtigen, der auf dem Wettbewerb von zwei Mannschaften unter Einhaltung der gleichen Regeln um den Sieg beruht (vgl. Lames, 1991, S.33). Hohmann und Lames (2005) ergänzen, dass die Selbstorganisation des Spielverhaltens, der hohe Zufallsanteil und die Ungewissheit des Spielausgangs aufgrund der Nicht-Linearität zwischen Spielfähigkeit, Spielleistung und Spielerfolg, sowie der Null-Summen Charakter des Spielerfolgs, welcher im Wesentlichen beschreibt, dass die Leistung im Wettkampf nur relativ zum Gegner erbracht werden kann, zusätzlich zu berücksichtigen seien (vgl. Hohmann & Lames, 2005, S.133ff.).

Siegle (2013) vereinfacht darauf aufbauend die Komplexität der Interaktionsprozesse im dynamischen System Fußball, indem er dieses aus einer systemtheoretischen Perspektive zergliedert. Das System Fußballspiel (System AB) bestehe aus zwei Mannschaften (System A und System B), die gegeneinander spielen. Jede Mannschaft bestehe aus Subsystemen, d.h.

aus Positionsgruppen (Systeme A1, A2, A3 und Systeme B1, B2, B3) und Einzelspielern (Systeme A1a-A3b und Systeme B1a-B3b). Beide Mannschaften versuchen, ihre jeweiligen Spielziele zu erreichen, d. h. ein Tor zu erzielen und die gegnerische Mannschaft an einem Torerfolg zu hindern. Die Wichtigkeit dieser beiden Attraktoren werde je nach Subsystemebene unterschiedlich gewichtet. So hätten beispielsweise die Verteidiger die Priorität, die gegnerische Mannschaft an einem Torerfolg zu hindern, während die Stürmer die Priorität hätten, ein Tor zu erzielen. Alle einzelnen Spieler verfolgten jedoch in gewissem Maße beide Ziele (vgl. Siegle, 2013, S. 23f.). Der Autor argumentiert, dass ein Fußballspiel trotz seiner Komplexität eine scheinbar organisierte Struktur aufweise. Er schlussfolgert, dass die Gründe dafür im System selbst liegen müssten, da es sich um ein weitgehend ein geschlossenes System handle und benennt zwei mögliche: Die teamübergreifende und die teaminterne Systemkopplung. Bei ersterer handle es sich um unterschiedlich starke Kopplungen und Interaktionen zwischen den Positionsgruppen und einzelnen Spielern beider Mannschaften, während sich letztere auf Interaktionen innerhalb der Positionsgruppen und einzelnen Spieler einer Mannschaft beziehen (vgl. ebd. S.25f.). Diese logische und leicht nachvollziehbare Struktur ist in Abbildung 3 dargestellt.

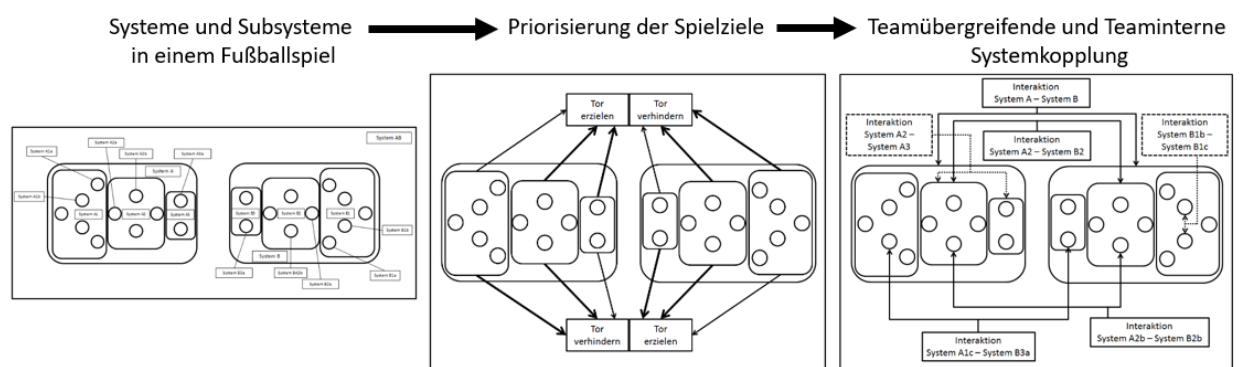


Abbildung 3: Systemtheoretische Gliederung des Fußballspiels. Quelle: Siegle, 2013, S. 23,24,26

Schmalhofer (2015) beschäftigt sich mit der Stabilität und Störungen des dynamischen Systems Fußballs und spricht in diesem Zusammenhang von einer grundsätzlichen Metastabilität, die in torkritischen Situationen in einen instabilen Zustand versetzt werde. Metastabilität ist definiert als ein Zustand, der bei kleinen Veränderungen stabil, bei größeren Abweichungen aber instabil ist. Im Fußball sei die Metastabilität durch den ständigen Wechsel zwischen offensivem Verhalten in Ballbesitz und defensivem Verhalten bei gegnerischem gekennzeichnet, mit dem sich im Phasenübergang auch das jeweilige Spielziel ändere. Die ständige Bewegung der Spieler und die ständig wechselnde Grundordnung der Mannschaften führt zu kleinen Veränderungen, die das System von einem stabilen in einen instabilen Zustand versetzen können. Diese

Verschiebungen werden durch Perturbationen verursacht, die als Störungen der Ordnung definiert werden, die durch erfolgreiche Offensivaktionen oder defensive Fehler verursacht werden und zu torkritischen Situationen führen (vgl. Schmalhofer, 2015, S. 19ff.).

Die beschriebenen Ansätze der dynamischen Systemtheorie bieten eine Möglichkeit, die Komplexität im Sportspiel zu analysieren. Hierbei geht es weniger darum, eine bestimmte Anzahl an Einzelaktionen zu zählen, die die Komplexität zwar auf ein messbares Niveau reduziert, aber die ganzheitliche sportliche Leistung vernachlässigen, sondern vielmehr um das tatsächliche Spielverhalten als Ganzes. Entscheidend ist es dabei Momente im Spielverlauf zu identifizieren, die eine Veränderung auslösen und das dynamische System in die Instabilität führen (vgl. ebd. S. 78ff).

Um solche Momente systematisch zu erfassen und quantifizieren zu können, bietet sich die Methode der systematischen Spielbeobachtung an. Diese basiert auf den Überlegungen von Lames (1994), der innerhalb der Sportspielbeobachtungsverfahren zwischen der systematischen Spielbeobachtung, dem Scouting und der subjektiven Eindrucksanalyse unterscheidet (vgl. Lames, 1994, S. 23f.). Die subjektive Eindrucksanalyse und die systematische Spielbeobachtung können nach Schmalhofer (2015) als Pole eines Kontinuums betrachtet werden, welche die optimale Beschreibung von Sportspielen zum Ziel haben. Während die subjektive Eindrucksanalyse flexible, nicht schriftlich fixierte Merkmale verwendet, basiert die systematische Spielbeobachtung auf genau festgelegten und fixierten Merkmalen. Der Autor stellt fest, dass sich beide Methoden in ihren Stärken und Schwächen ergänzen, da sie auf unterschiedlichen Ebenen agieren. Die Stärken der Systematischen Spielbeobachtung liegen in ihrer Systematik und Objektivität, die es ermöglicht, ein exaktes Abbild des Spielgeschehens zu erstellen, differenzierte Analysen durchzuführen und den Spielverlauf zu rekonstruieren (vgl. Schmalhofer, 2015, S. 80).

Die systematische Spielanalyse als Ganzes - unabhängig davon, ob sie qualitativer oder quantitativer Natur ist – dient deshalb im Kontext der trainingswissenschaftlichen Gegenstandsbereiche als Hilfsmittel zur Bewertung der sportlichen Leistungsfähigkeit eines einzelnen Spielers oder einer gesamten Mannschaft und befindet sich somit an der Schnittstelle zwischen Training und Wettkampf. Im Fokus steht hierbei die Messung bzw. Bewertung der Effektivität individueller und kollektiver Spielhandlungen, die wiederum für die Steuerung und inhaltliche Gestaltung des Trainings genutzt werden kann.

Scouting wurde von Lames (1994) ursprünglich als „Beobachtungsmethode zum Ausspähen einer gegnerischen Mannschaft“ (Lames, 1994, S. 23) definiert, und aufgrund fester und flexibler Merkmale sowie der teilweisen schriftlichen Fixierung der Eindrücke und Beobachtungen in der Mitte des Kontinuums angesiedelt (vgl. ebd., S. 24). Im Rahmen dieser Arbeit ist der

Scoutingbegriff dahingegen als systematische, quantitative (d.h. einzig auf Daten basierende) Einzelspielerbeobachtung zu verstehen, die vor Transfers durchgeführt wird.

Die wichtigsten Anwendungsbereiche von Data Analytics in der Spielanalyse, im Work-Load-Management und im Scouting wurden bereits in Abschnitt 2.2 ausführlich behandelt (vgl. 2.2). An dieser Stelle erscheint deshalb eine zusammenfassende Einordnung dieser Anwendungsbereiche in den Prozess der Generierung von Trainingszielen aus dem Wettkampfverhalten sinnvoll. Diese Einordnung wird in Abbildung 4 veranschaulicht.

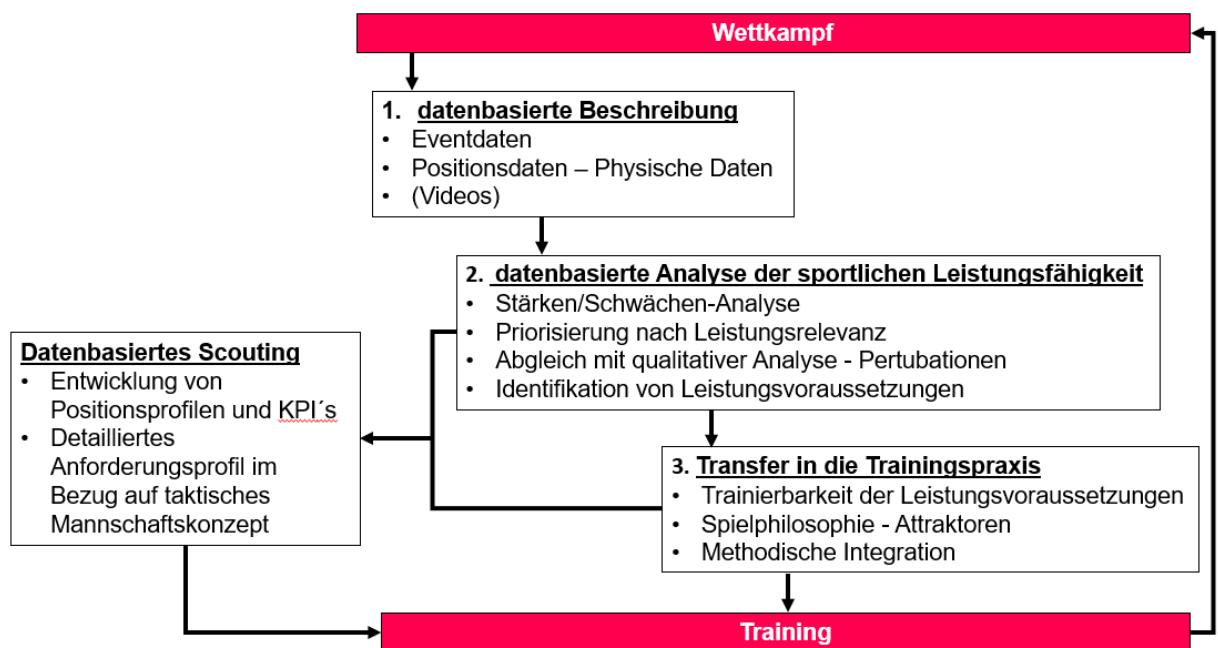


Abbildung 4: Anwendungsgebiete von Data Analytics im Prozess der Generierung von Trainingszielen aus Wettkampfverhalten. Quelle: eigene Darstellung (modifiziert nach Lames, 1994, S.30)

Aus Gründen der Übersichtlichkeit wurde in der Darstellung im Bereich der Spielanalyse bewusst auf die Integration der datenbasierten Pre-Match-Analyse des nächsten Gegners verzichtet, obwohl diese in der Praxis ebenfalls einen großen Stellenwert für die Trainingsinhalte hat. Zudem wurde auf eine detaillierte Darstellung der Spielvorbereitungsphase, bei der im Wesentlichen Elemente des datenbasierten Workload-Managements eine Rolle spielen, verzichtet. Dies erfolgte zugunsten der systematischen Integration des Scoutings, welches den Kernbereich dieser Arbeit darstellt.

Das Scouting gilt aus der Perspektive der Data Analytics als eigenständiges Anwendungsgebiet und erfordert aus trainingswissenschaftlicher Sicht eine andere Beobachtungsmethode. Dennoch ist es für Vereine mit klarer Spielphilosophie und eindeutigen Prinzipien in den verschiedenen Spielphasen zu empfehlen, die individuellen Positionsprofile, vereinspezifischen KPIs und notwendigen Leistungsvoraussetzungen potenzieller Neuzugänge im Hinblick die Anforderungen



des taktischen Mannschaftskonzepts, direkt aus der datenbasierten Analyse der eigenen Mannschaft abzuleiten. Das Wissen über den Grad der Trainierbarkeit der erforderlichen Leistungsvoraussetzungen ist für das datenbasierte Scouting ebenfalls von zentraler Bedeutung. Daher bietet die Integration des Scoutings in die Darstellung des Prozesses der Generierung von Trainingszielen aus Wettkampfverhalten einen erheblichen Mehrwert für den Leser, indem sie ein besseres Verständnis für die Wechselwirkungen und gegenseitigen Abhängigkeiten der verschiedenen Anwendungsgebiete von Data Analytics in der Fußball-Praxis vermittelt, welcher im Vergleich zu einer vermeintlichen Vermischung der trainingswissenschaftlichen Teilbereiche überwiegt.

### **3.2 Datenarten und deren Informationsgehalt**

Prinzipiell kann man im Kontext des Fußballs zwischen sportlichen Daten und außersportlichen Daten mit einem indirekten Bezug zum Sport unterscheiden. Eine weitere Differenzierung innerhalb der sportlichen Daten ergibt sich im Hinblick auf die Erhebungssituation, wobei zwischen im Spiel und im Training erhobenen sportlichen Daten differenziert wird.

Die Event-/Ereignisdaten sowie Positions-/Trackingdaten, welche auch als Grundlage für die Ableitung von Geschwindigkeitsdaten dienen können, befinden sich unabhängig vom spezifischen Anwendungsgebiet im Zentrum des Interesses. Eventdaten beschreiben systematisch erfasste, spezifische und präzise definierte Ereignisse, wie beispielsweise Pässe oder Schüsse, welche von geschultem Personal manuell erfasst werden. Die DFL greift hierbei auf einen speziellen Spieldatenkatalog zurück, der etwa 30 Ereignisse mit über 100 Attributen definiert. Andere Datenprovider verwenden hingegen individuelle Definitionskataloge. Allerdings werden bei der Erfassung von Eventdaten lediglich der jeweilige Spieler und dessen Aktionen am Ball berücksichtigt, während die übrigen 21 Spieler nicht erfasst werden. Deshalb besitzen Eventdaten bei isolierter Betrachtung nur einen begrenzten Mehrwert. Im Gegensatz dazu beschreiben Positions-/Trackingdaten die Position jedes Spielers, des Schiedsrichters sowie des Balls (in Form von x/y/z-Koordinaten) zu 25 Zeitpunkten pro Sekunde, welche u.a. durch kameragestützte Trackingsysteme erfasst werden. Auf die unterschiedlichen Datenerhebungsmethoden wird in 3.3 detaillierter eingegangen. Es sollte angemerkt werden, dass Video-Daten, aus denen durch Tagging Event- und Positionsdaten auch selbstständig extrahiert werden können, in diesem Kontext von Bedeutung sind.

Durch die Kombination von Event- und Positionsdaten können sämtliche Aktionen und Bewegungen aller Spieler erfasst und in Zusammenhang gesetzt werden, woraus sich präzise 2D-Modelle sowie KPI's ableiten lassen. Jedes Ereignis wird dabei mit seinen (x,y)-Koordinaten erfasst, was für die Datenanalyse äußerst wertvoll ist, da neben den quantitativen Aspekt die Bedeutung jeder Aktion im Spielkontext miteinbezogen werden kann. Mit darauf aufbauenden

Berechnungen von Modellen und Metriken wie dem Expected Goals (xG)- Wert, die auch als Advanced Metrics bezeichnet werden, wird ein erheblicher Mehrwert geschaffen. Diese Metriken ermöglichen es, die Leistungen von Einzelspielern und Mannschaften vergleichbarer zu machen, indem sie neben Anzahl und Qualität der Aktionen auch den Spielsituationskontext berücksichtigen. Neben den Expected Goals und seinen verschiedenen Ableitungen werden inzwischen auf individueller und Mannschaftsebene viele weitere Metriken wie beispielsweise Expected Assists (xA), Expected Threat (xT), Passes Allowed per Defensive Action (PPDA) oder dem Packing erhoben.

Die Erfassung von Positionsdaten ist heutzutage sowohl im Training als auch im Spiel allgemein etabliert. Im Gegensatz dazu wird die umfassende Erfassung von Eventdaten im Training aufgrund ökonomischer Faktoren weniger häufig praktiziert.

Physischen Daten bilden eine weitere Schnittstelle zwischen Spiel und Training. Diese werden nicht in den offiziellen Spieldaten erfasst. Das Regelwerk erlaubt jedoch mittlerweile die Nutzung von elektronischen Transpondern, Wearables und Tracking-Sensoren, die im Training bereits seit langer Zeit genutzt werden und zusätzlich interne Belastungsdaten wie beispielsweise die Herzfrequenz messen. Diese Daten vervollständigen in Kombination mit anderen Tracking-Daten das Gesamtbild des Belastungsprofils eines Spielers und stehen zumindest in Bezug auf die eigene Mannschaft zur Verfügung, sofern sich der Verein zur Nutzung im Spiel entscheidet. Zusätzlich zu den live erhobenen Daten werden die physischen Daten durch interne Leistungsdiagnostik-Daten ergänzt. Die physischen Daten sind im Anwendungsgebiet des Workloadmanagements eng mit medizinischen Daten verbunden. Diese beinhalten beispielsweise Creatinkinase- und Laktatwerte, die individuelle Anaerobe Schwelle die Schlafqualität, sowie Verletzungshistorien und Therapieverläufe.

Im Rahmen der offiziellen Spieldaten werden historische Daten bereitgestellt, die Stamm-, Spielinformations-, Event- und Positionsdaten aus vergangenen Spielzeiten enthalten. Um ein nachhaltiges Arbeiten im Bereich der Belastungssteuerung zu gewährleisten, sollten vergleichbare historische Daten auch für das Training in der vereinseigenen Datenbank verfügbar sein. Stammdaten beinhalten Informationen, die in der Regel nur geringen Veränderungen unterliegen, wie beispielsweise Stadien, Spieler, Teams und Spielpläne. Des Weiteren werden Spielinformationsdaten bereitgestellt, welche Informationen über das Spiel selbst wie beispielsweise Aufstellungen, Umweltbedingungen und Schiedsrichter beinhalten. VAR-Daten stehen ebenfalls zur Verfügung, sind aber für alle Anwendungsbereiche von Data Analytics von untergeordneter Bedeutung.

Im Rahmen des Scouting-Prozesses können sämtliche Spieldaten von unterschiedlichen Datenanbietern genutzt werden, wobei mögliche Unterschiede in der Definition einzelner Events

zu berücksichtigen sind. Hierbei liegt der Schwerpunkt auf den Event- und Positionsdaten sowie den daraus abgeleiteten Metriken. Die daraus gewonnenen Informationen werden durch Spielerbewertungen von Live-Scouts und/oder der qualitativen Analyse im Video-Scouting ergänzt. Des Weiteren stellen Transfermarkt-Daten wie die Vertragslaufzeit, das Gehalt oder die Marktwertentwicklung eine entscheidende Komponente im Scouting dar. Bei der Entscheidung über die Verpflichtung eines Spielers kann auch das Vermarktungspotenzial eine Rolle spielen, welches durch die Analyse von Social-Media-Daten und Erwähnungen in Artikeln ermittelt werden kann. In Abbildung 5 sind die verschiedenen Datenarten schematisch zusammengefasst.

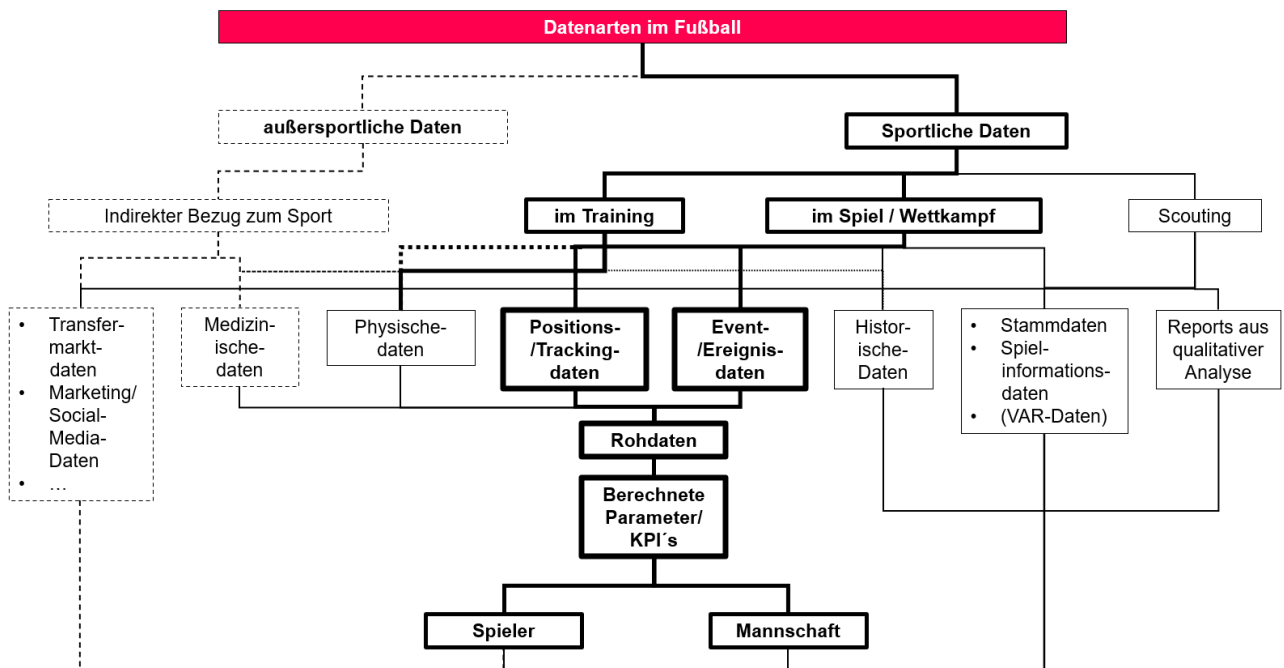


Abbildung 5: Datenarten im Fußball Quelle: eigene Darstellung

### 3.3 Datenerhebungsmethoden, deren Eigenschaften und Qualitätsmerkmale

Allein im Rahmen der offiziellen Spieldaten werden für jedes Spiel in der 1. und 2. Bundesliga etwa 3,6 Millionen Positionsdatenpunkte und 1600 Spielereignisse erfasst (vgl. DFL, 2020). Während die Eventdatenerfassung noch überwiegend manuell oder teilweise automatisiert mit manueller Nachbearbeitung erfolgt, können Positionsdaten je nach Art der Datenerfassung teilweise oder vollständig automatisch erfasst werden. In der technischen Umsetzung lassen sich grundsätzlich drei verschiedene Verfahren zur Positionsdatenerfassung unterscheiden: optisches Tracking (OT), GPS-basierte Verfahren und radarbasierte Datenerfassung (Local Positioning Measurement, LPM). Jedes dieser Verfahren hat unterschiedliche technische Anforderungen und

Ansätze, die in der praktischen Anwendung Vor- und Nachteile aufweisen. Diese werden im Folgenden kurz erläutert.

Die Erfassung von Eventdaten erfolgt anhand eines definierten Eventdatenkatalogs (vgl. 3.2). Während jedes Spiels sind vier Mitarbeiter für die offizielle Erfassung der Spieldaten zuständig. Der Speaker, welcher sich stets im Stadion aufhält, beschreibt das Live-Spiel in einer Codesprache dem Writer, welcher im Sportec Solutions Zentrum sitzt. Der Writer gibt die Ereignisse in Echtzeit in das System ein. Zwei weitere Mitarbeiter, die Operater, überprüfen die eingegebenen Eventdaten anhand verschiedener Live-Videoaufnahmen, um diese zu genehmigen, zu korrigieren oder gegebenenfalls zusätzliche Informationen hinzuzufügen (vgl. DFL, 2023). Andere Anbieter wie z.B. Stats Perform, Statsbomb, Wyscout oder Sportradar erfassen Daten auf ähnliche Weise, jedoch setzen nicht alle auf eine direkte Erfassung der Daten aus dem Stadion, sondern greifen teilweise auch auf TV-Bilder oder andere Videoquellen zurück. Um die Validität der Daten zu gewährleisten, ist neben der Unterscheidung der Rechte (offiziell vs. inoffiziell) vor allem die detaillierte Beschreibung der einzelnen Ereignisse im jeweiligen Definitionskatalog zu beachten. Es kann vorkommen, dass bestimmte Ereignisse durch ihre Definition enger oder weiter gefasst sind, als es ein rein praktisches Verständnis des Fußballs vermuten lassen würde. Mögliche doppelte Erfassungen der gleichen Spielaktion sind ebenfalls zu berücksichtigen, insbesondere wenn es sich dabei um ein Attribut eines Ereignisses handelt.

Die manuelle Datenerfassung bleibt trotz Fortschritten in (teil-)automatisierten, vorwiegend videobasierten Bilderkennungssystemen ein entscheidender Prozess, um eine angemessene Datenqualität sicherzustellen. Eine vollständig automatisierte Erfassung ist in den meisten Fällen zum jetzigen Zeitpunkt noch nicht möglich, da menschliche Kontrolle erforderlich ist. Trotz detaillierter Definitionskataloge besteht bei der Beteiligung von Menschen immer ein subjektiver Interpretationsspielraum und die Gefahr von Wahrnehmungsfehler, die zu Ungenauigkeiten führen können. Jedoch werden diese Fehler durch verschiedene Kontrollinstanzen minimiert. Zudem sollte der Zeitverzug berücksichtigt werden, der durch diesen Prozess entsteht. Obwohl das Ereignis zeitnah zur Live-Aktion erfasst und in den ersten Instanzen korrigiert wird, erfolgt die endgültige Freigabe zumeist erst nach dem Spiel.

Für die Erhebung von Positionsdaten im Rahmen der offiziellen Spieldaten wird optisches (Kamera-)Tracking verwendet. In der ersten und zweiten Bundesliga sind je nach Austragungsort zwischen 16 und 20 Kameras an verschiedenen Stellen im Stadion positioniert, um positionsbasierte Daten während der Spiele zu erfassen. Die Kameras ermöglichen die Verfolgung einzelner Spieler in komplexen Spielsituationen wie Eckbällen oder Freistößen. Durch den Einsatz fortschrittlicher Machine-Learning-Algorithmen und künstlicher Intelligenz kann das Tracking-System neben den X-Y(-Z)-Daten (vgl. 3.2) eine Vielzahl von Statistiken ermitteln wie

z.B. Informationen wie Beschleunigungswerte für Spieler oder die Länge von Pässen (vgl. DFL, 2023).

Optische Tracking-Systeme, oft auch als Computer Vision bezeichnet, zur Positionsdatenerfassung haben den Vorteil, dass sie keine zusätzliche Sensorik an den Spielern benötigen. Die Systeme verwenden Methoden der Bildverarbeitung, um Spieler anhand von Kameraperspektiven zu identifizieren und ihre Laufwege nachzuvollziehen. Allerdings erfordern die meisten Systeme manuelle Interventionen bei Verdeckungen oder Verlust der Spielerpositionen. Sie haben Schwierigkeiten bei sich kreuzenden Wegen der Spieler oder wenn sie Jubeltrauben bilden. Zudem sind sie anfällig für Witterungsbedingungen oder unterschiedliche Beleuchtungen (z.B. Schattenwurf). Diese Erfassungsprobleme führen zur Notwendigkeit einer Nachbearbeitung der Daten (vgl. Memmert & Raabe, 2017, S. 65f). Perl (2017) ergänzt, dass automatisch erfasste Positionsdaten für Spieler in akzeptabler Qualität verfügbar seien, die Erfassung der Ballposition jedoch teilweise gravierende Fehler aufweise (vgl. Perl, 2017, S. 69f).

Barthel (2017) stellt unterschiedliche Ansätze zur Optimierung der automatisierten Erkennung und Verfolgung von Objekten in Videos vor. Einige dieser Methoden wurden bereits von verschiedenen Anbietern wie Stats Perform umgesetzt, welche zur Positionsdatenerfassung Computer Vision auf Basis von TV-Bildern nutzen. Allerdings soll an dieser Stelle nicht detaillierter darauf eingegangen werden (vgl. Barthel, 2017, S.11ff). Es erscheint ausreichend zu betonen, dass Verfahren des optischen Trackings, die auf TV-Bildern oder anderen Videosequenzen basieren, selbst bei Verbesserungen der Algorithmen und Methoden in Zukunft, die Genauigkeit der originalen Erfassung mit bis zu 20 Kameraperspektiven kaum erreichen können. Trotzdem sind die daraus gewonnenen Daten valide und bieten einen enormen Mehrwert hinsichtlich der Verfügbarkeit von Positionsdaten aus unterschiedlichen Ligen. Es ist jedoch nicht zielführend, die Daten verschiedener Anbieter im Scouting zu vergleichen. Stattdessen sollte man sich auf Daten desselben Anbieters beschränken, da davon ausgegangen werden kann, dass unabhängig von der Liga eine ähnliche Verzerrung vorliegt. In diesem Zusammenhang kann darüber hinaus auf Rahimian und Toka (2022) verwiesen werden, die eine detaillierte Klassifikation von konventionellen und Deep-Learning-Methoden des optischen Trackings präsentieren, die notwendigen Schritte zur Vorverarbeitung von Trackingdaten sowie die Herausforderungen in diesem Bereich erläutern und weiteren Forschungsbedarf aufzeigen.

Zur Erhebung von Positionsdaten im Sport können neben dem optischen Tracking auch GPS- und LPM-Verfahren eingesetzt werden. Diese ermöglichen nicht nur die Erfassung der Positionen der Spieler, sondern bieten auch die Möglichkeit, physiologische Daten wie z. B. die Herzfrequenz zu erfassen. Im Gegensatz zum optischen Tracking erfordern diese Verfahren jedoch eine Sensorik an den Spielern.

Die Positionsdatenerfassung mittels GPS-Systemen beruht auf dem gleichnamigen, vom US-Militär entwickelten, Navigationssatellitensystem. Zur Präzisierung der Erfassung der Spielerlaufwege werden zusätzlich Accelerometer, Gyroskope und Kompass verwendet. Die Tracker, die darüber hinaus üblicherweise physiologische Daten erfassen, zeichnen sich durch ihre einfache Handhabung aus, da nur ein Transponder pro Spieler erforderlich ist, um die Daten zu sammeln, und ein Empfangs- sowie ein Stromversorgungsgerät. Die Anwendung der Technologie ist für Profivereine daher sowohl kosteneffizient als auch logistisch praktisch. Die Technologie hat jedoch auch Einschränkungen, insbesondere in Bezug auf die Genauigkeit der Positionsbestimmung auf dem Spielfeld, die mit Hilfe von differenzialem GPS als lokale Referenz verbessert werden kann, sowie auf die Funktionalität nur im Freien. Trotzdem sind GPS-Systeme insbesondere im Anwendungsgebiet des Work-Load-Managements ein wertvolles Instrument zur Datenerfassung (vgl. Memmert & Raabe, 2017, S. 63ff).

Systeme, die auf Radar- und Mikrowellentechnologie basieren und als LPM-Systeme bekannt sind, stellen eine dritte Möglichkeit zur Positionsdatenerfassung dar. Ähnlich wie bei GPS tragen Spieler kleine Transpondereinheiten. Der Unterschied besteht darin, dass die Positionsbestimmung nicht über ein Satellitennavigationssystem erfolgt, sondern über mehrere fest installierte Empfängereinheiten, die rund um das Spielfeld positioniert sind. Wenn das Signal eines Spielers einen Transponder erreicht, wird es zurückgeschickt und von den Empfängereinheiten aufgefangen, wobei je nach Position des Spielers unterschiedliche Signallaufzeiten entstehen. Ein zentraler Server berechnet dann mithilfe der Triangulation die genaue Position des Spielers und liefert diese Daten in Echtzeit an einen Laptop, auf dem sie von Analysten ausgewertet werden können. Im Gegensatz zu GPS-Systemen, die Satelliten verwenden und in Innenräumen nicht funktionieren, sind LPM-Systeme auch in geschlossenen Umgebungen einsetzbar. Allerdings erfordert die Installation dieser Systeme im Stadion oder auf dem Trainingsplatz eine aufwendige Kalibrierung, die Montage von Empfängereinheiten am Stadionsdach oder Flutlichtmasten und ist verhältnismäßig teuer. Während dies im eigenen Stadion und den eigenen Trainingsplatz machbar ist, ist es bei Auswärtsspielen oder Trainingslagern deutlich aufwändiger (vgl. ebd., 2017, S. 67f).

Bastida-Castillo et al. (2019) haben die Genauigkeit von GPS- und LPM-basierten Systemen verglichen und deren Eignung für taktische Analysen untersucht. Die Ergebnisse zeigten, dass das LPM-System im Vergleich zum GPS-System eine signifikant höhere Genauigkeit aufwies. Die durchschnittliche absolute Fehlerquote für die X-Y-Koordinaten betrug für GPS  $41,23 \pm 17,31$  cm bzw.  $47,6 \pm 8,97$  cm und für LPM  $9,57 \pm 2,66$  cm bzw.  $7,15 \pm 2,62$  cm. In "small sided Games", d.h. 4 gegen 4 bis 7 gegen 7 auf reduziertem Spielfeld, ergab sich zwischen den beiden Systemen hinsichtlich der taktischen Variablen ein Unterschied von bis zu 8,31%. Beide Systeme zeigten insgesamt eine gute Test-Reliabilität und interne Zuverlässigkeit, jedoch eignet sich das

LPM-System aufgrund der höheren Genauigkeit besser für taktische Analysen (vgl. Bastida-Castillo et al., 2019, S.5ff).

Die Wahl des Systems zur Positionsdatenbestimmung hängt neben finanziellen Ressourcen primär davon ab, ob physische oder taktische Leistungsüberwachung im Fokus steht. Bei der Analyse der Daten muss berücksichtigt werden, dass die Vergleichbarkeit der mittels unterschiedlicher Verfahren erhobenen Daten nicht zwangsläufig gegeben ist. Insbesondere dann, wenn unterschiedliche Verfahren im Training und im (Auswärts-)Spiel verwendet werden. Darüber hinaus ist entscheidend, dass eine zuverlässige Erfassung der Ballposition, die am zuverlässigsten mit einem Sensor im Ball zu erreichen ist, gewährleistet ist.

## **4 Anforderungsprofil des Torspielers**

Im folgenden Kapitel wird zunächst die historische Entwicklung des Torspielers aus qualitativer Perspektive betrachtet. Es werden zudem die Ergebnisse wichtiger Analysen und Studien zur Anzahl von Torspieleraktionen pro Spiel zusammengefasst, welche als Grundlage für die Erstellung des Anforderungsprofils eines Torspielers im Wettkampf dienen. Anhand dieses Anforderungsprofils erfolgt eine Systematisierung der Event- und Positionsdaten sowie der Advanced Metrics, die von führenden Anbietern bereitgestellt werden. Diese Systematisierung bildet die Grundlage für eine systematische und unabhängige Datenanalyse im Torspieler-Scouting.

### **4.1 Vom Torspieler 1.0 zum Torspieler 4.0**

Im Jahr 1992 wurde die Struktur des Fußballspiels durch die Einführung der Rückpassregel wesentlich verändert. Diese Regel hatte insbesondere Auswirkungen auf das Anforderungsprofil des Torhüters, da es ihm fortan untersagt war, einen ihm als Rückpass zugespielten Ball mit den Händen aufzunehmen. Dies war sowohl für die Torhüter in Bezug auf Ballkontrolle und Passtechnik als auch für deren Mitspieler, die gezwungen waren, kreative Lösungen im Zusammenspiel mit dem Torhüter zu entwickeln, um sich aus gegnerischen Pressing-Situationen zu befreien, eine enorme Herausforderung. Es ist mittlerweile zunehmend anerkannt, dass die Kernaufgabe der Torhüter nicht ausschließlich darin besteht, Gegentore zu verhindern, sondern diese auch eine wichtige Rolle im Mitspielverhalten einnehmen. Aus diesem Grund erscheint im Kontext dieser Arbeit die Verwendung des Begriffs "Torspieler", der das aktuelle Anforderungsprofil besser beschreibt, angemessener.

Thaler (2019) erläutert die historische Entwicklung vom Torspieler 1.0 bis zum Torspieler 4.0. Diese Entwicklungsstufen sind von hoher Relevanz für das im Rahmen dieser Studie durchgeführte Clustering-Verfahren, welches als Basis für ein datenbasiertes Torspieler-Scouting dient. Die Torspieler der ersten Generation waren von der Einführung der Rückpassregel

überrascht und wurden nur in absoluten Notsituationen angespielt. Einige Torspieler dieser Zeit wiesen ihre Mitspieler sogar an, den Ball besser ins Aus zu spielen. Da die meisten Mannschaften zu dieser Zeit noch mit einem Libero agierten, mussten die Torspieler kaum in der Tiefe absichern. Die Einführung der Viererkette in den 2000er Jahren führte zur Entwicklung des Torspielers 2.0, der zunehmend als Absicherung bei gegnerischem Ballbesitz eingesetzt wurde und im eigenen Ballbesitz hauptsächlich dazu diente, den Ballbesitz zu sichern. Im Laufe der 2010er-Jahre etablierte sich die bis heute gültige Spielauffassung, in der es selbstverständlich ist, den Torspieler 3.0 ins Spiel zu integrieren. Die Unterstützung der Abwehrkette im Ballgewinnspiel (Mitspielverhalten bei gegnerischem Ballbesitz) steht dabei im Zentrum und erfüllt die Spielanforderungen bereits nahezu perfekt. Im eigenen Ballbesitz agiert der Torspieler 3.0 meist als Backup für die Feldspieler und bleibt im Angriffsspiel eher im Hintergrund. Seit Ende der 2010er-Jahre ist jedoch vereinzelt zu beobachten, dass der Torspieler in der Spieleröffnung eine zentrale Rolle einnimmt. Seine Einbindung als elfter Feldspieler und die damit einhergehende Schaffung von relativen Überzahlsituationen stellen die Entwicklungsstufe zum Torspieler 4.0 dar. Beispiele dafür sind die Umsetzung des FC Barcelona mit Marc-Andre ter Stegen, von Manchester City mit Ederson oder aus deutscher Perspektive Mannschaften von Christian Titz oder Tim Walter (vgl. Thaler, 2019, S. 42f).

Die vorliegende Beschreibung des Entwicklungsprozesses basiert auf qualitativer Spielanalyse und kann durch quantitative Analysen verschiedener Autoren, welche mittels der Anzahl verschiedener Aktionen pro Spiel ein Anforderungsprofil erstellten, trotz verhältnismäßig geringer Anzahl aussagekräftiger Analysen wissenschaftlich mehr als ausreichend fundiert werden. Aufgrund der unterschiedlichen Aufbereitung der Daten gestaltet sich die Ableitung von präzisen Vergleichen jedoch herausfordernd. Nichtsdestotrotz lässt sich übereinstimmend feststellen, dass die mit Abstand häufigsten Aktionen dem Mitspielverhalten im eigenen Ballbesitz zuzuordnen sind (vgl. Gruber & Zeilinger, 2021, S.19). Im Folgenden werden die bedeutsamsten Ergebnisse in Hinblick auf die Kategorisierung der individuellen Aktionen, welche für diese Arbeit relevant ist, prägnant zusammengefasst.

Loy (1991) lieferte einen ersten Versuch, das Torwartspiel hinsichtlich seiner Anforderungen im Fußball wissenschaftlich zu fundieren, unterschied jedoch lediglich zwischen Offensiv- und Defensivaktionen (vgl. Krebs, 2008, S. 10). Krebs (2008) hat die Spielaktionen von Torspielern in vier Hauptkategorien unterteilt:

1. Mitspielende Aktionen: Hierbei handelt es sich um Rückpässe, das Ablaufen von Steilpässen und das Abfangen von Flanken.
2. Abwehrende Aktionen: Aktionen, die ohne das Eingreifen des Torspielers zum Gegentor geführt hätten, ähnlich der Zielverteidigung.



3. Standardaktionen: ruhende Bälle nach Spielunterbrechungen wie Freistöße und Abstöße.
4. Spieleröffnungen: Aktionen, die der Torspieler ausführt, nachdem er den Ball in die Hand genommen hat, wie Abwürfe, Abschlüsse und Flugbälle.

Die Torspieleraktionen werden darüber hinaus in Hinblick auf das eingesetzte Körperteil, Hand bzw. Fuß, analysiert. Außerdem werden die Aktionsräume der Torspieler untersucht und in die Kategorien "innerhalb" oder "außerhalb" des Strafraumes unterschieden (vgl. ebd. S.19ff). In der BL-Saison 2007/2008 griff ein Torspieler durchschnittlich etwa 40-mal pro Spiel aktiv ins Spielgeschehen ein. Von insgesamt 9.139 analysierten Aktionen wurden 38% mit der Hand und 62% mit dem Fuß ausgeführt. Die Mitspielenden Aktionen machen mit 39% den größten Anteil der Hauptaktionen aus, während abwehrende Aktionen mit 9% den geringsten Anteil hatten (vgl. ebd. S.28). Eine Analyse der EM-Spiele 2008 lieferte ähnliche Ergebnisse (vgl. ebd. S. 60). Nach Aufbereitung der Daten aus der Loy-Studie stellt Krebs im Vergleich einige Veränderungen fest, die auf die Einführung der Rückpassregel zurückzuführen sind. Darunter fallen eine deutliche Abnahme der Spieleröffnungen mit der Hand, weniger abwehrende Aktionen und deutlich mehr Mitspielende Aktionen (vgl. ebd. S.53.ff).

Memmert und Rechner (2010) haben darauf aufbauend eine Weltstandsanalyse durchgeführt, die auf einer quantitativen Spielanalyse von 40 Spielen aus verschiedenen Wettbewerben basiert. Dabei haben sie die Generalisierbarkeit, Tiefenstruktur (Gründe für Gegentore) und Auswertbarkeit verbessert, indem sie nicht nur die Ballkontakte, sondern auch die einzelnen Aktionen gezählt haben. Die Autoren unterscheiden zwischen:

1. Torverteidigung: direkte Torverteidigung
2. Raumverteidigung: indirekte Torverteidigung
3. Spieleröffnung: Offensivaktion bei eigenem Ballbesitz (inkl. Standards und Spieleröffnungen mit der Hand)

Im Bereich der Torverteidigung unterscheiden die Autoren zwischen Grundtechniken, Abdrücken, Eins gegen Eins-Situationen und Ballaufnahme, während im Bereich der Raumverteidigung zwischen Flanken und Mitspielen d.h. indirekter Torverteidigung im und um den Strafraum differenziert wird. Die Spieleröffnung macht mit 73,19% der Aktionen den größten Anteil aus, während die Tor- und die Raumverteidigung 16,88% bzw. 7,39% der Aktionen ausmachen. Interessanterweise wurden bei 25% aller Gegentore Torspielerfehler identifiziert. Die drei Hauptursachen von Fehlern seien das Treffen einer falschen Entscheidung, das Stellungsspiel und die Grundposition. Obwohl die Bereiche der Spieleröffnung und Grundtechniken den größten Anteil an Aktionen ausmachen (90,7%), traten hierbei kaum Fehler auf und sie werden daher als nicht spielentscheidend angesehen. Im Gegensatz dazu wurden Fehler hauptsächlich in den Bereichen Abdrücke, Eins gegen Eins und Flanken und Mitspielen gemacht, die zwar nur wenige

Aktionen (1%, 1,28%, 7,39%) ausmachen, aber als spielentscheidende Defensivtechniken identifiziert wurden, da 93,45% aller Fehler in diesen Bereichen auftreten (vgl. Memmert & Rechner, 2010).

Ziegler (2021) liefert mit der Analyse der EM 2020 die aktuellsten Daten bezüglich der Torspieler-Aktionen. Von diesen wurden knapp 73% als Offensivaktionen mit dem Fuß oder der Hand kategorisiert, während sich das verbleibende Viertel fast gleichmäßig auf die Raumverteidigung (14%) und Zielverteidigung (13%) aufteilte. Der Autor differenziert darüber hinaus zwischen Aktionen, die einen höheren Anspruch an den Torspieler haben, wie z.B. Zeit-, Raum-, Gegner- und Präzisionsdruck, und solchen ohne Anspruch. Ohne die "Aktionen ohne Anspruch" nehme die Zielverteidigung mit 40% einen signifikant höheren Anteil ein, was die Bedeutung des Torhüters betone. Obwohl diese Differenzierung sinnvoll erscheint, ist die Schlussfolgerung fragwürdig, da sie auf einer Verzerrung der ursprünglichen Daten beruht. Ziegler identifiziert lediglich 44% der Gegentore als unhaltbar, während bei 40% der Gegentore eine Beteiligung des Torspielers nachzuweisen sei und 16% als haltbar eingestuft wurden, ohne jedoch eine genaue Abgrenzung dieser Beteiligung zu liefern. Die Abstimmung zwischen Torspieler und Mannschaft wird als wesentliche Fehlerquelle genannt. Von den durchschnittlich 2,2 Raumverteidigungsaktionen pro Spiel, die aus 1,7 Spielaktionen und 0,5 Standardsituationen bestehen, stehen die Abwehr von Flugbällen und Hereingaben von außen im Mittelpunkt (43%), gefolgt von Pässen in die Tiefe (15%) und der Querpassverteidigung (12%). Von den 73% der Offensivaktionen wurden mehr als die Hälfte als "anspruchsvolle" Aktionen kategorisiert. Von diesen machten rund 71% die Spielfortsetzung nach einem Rückpass aus, während 11% die Spielfortsetzung nach einer Defensivaktion waren. Die Spieleröffnung machte 14% der Aktionen aus, während 4% dem Umschaltverhalten zuzuschreiben waren (vgl. Ziegler, 2021).

Gruber und Zeilinger (2021) vergleichen die Ergebnisse von Krebs und Memmert/Rechner mit einer von Ziegler erstellten Analyse der WM 2018. Auf eine detaillierte Darstellung von Zieglers WM-Analyse wird bewusst verzichtet, da diese fast keine nennenswert neuen Erkenntnisse liefert. Bemerkenswert ist lediglich, dass der Großteil der Fehler in der Zielverteidigung (78%) und der Raumverteidigung (87%) auf taktische Fehler zurückzuführen war. Die Autoren erkennen in diesem Vergleich zwischen den Jahren 2010 und 2018 einen Trend, wonach technische Fehler (Grundstellung und Technik) prozentual zu deutlich weniger Gegentoren führen. Dies wird durch die bessere technische Ausbildung der Torspieler erklärt (vgl. Gruber & Zeilinger, 2021, S. 21f).

Das Anforderungsprofil des Torspielers setzt sich nicht nur aus den bereits erwähnten technisch-taktischen Leistungsfaktoren, die anhand der Spieldaten messbar sind, zusammen, sondern beinhaltet auch konditionelle, physische/konstitutionelle und psychische/kognitive Faktoren. Im Kontext des datenbasierten Scoutings stellen diese oft leistungslimitierenden Faktoren eine

Herausforderung dar, da entsprechende Daten häufig nur begrenzt verfügbar sind, weil diese entweder nur in internen Leistungstests vom aktuellen Verein des potenziellen Neuzugangs erhoben werden oder eine valide Quantifizierung der Faktoren durch standardisierte sportwissenschaftliche Testverfahren schlicht nicht möglich ist. Es gibt jedoch Ausnahmen wie die Daten bezüglich der Schnelligkeit und Laufleistung, die aus Positionsdaten abgeleitet werden können und somit über offizielle Spieldaten verfügbar sind. Es ist jedoch zu beachten, dass die Genauigkeit dieser Werte stark vom jeweiligen Spielkontext abhängig ist und deshalb keine Aussage über die maximale Leistungsfähigkeit eines Spielers gemacht werden kann. Einfache antropometrische Daten wie die Körpergröße können eine weitere Ausnahme darstellen, sofern diese vom jeweiligen Verein als relevanter Faktor betrachtet werden. Im Hinblick auf das Ziel dieser Arbeit sind die genannten Leistungsfaktoren von untergeordneter Bedeutung, da die Datenverfügbarkeit unzureichend ist. Daher wird auf Krebs (2008, Kapitel 3) und Gruber/Zeilinger (2022, Kapitel 2 und 3) verwiesen, um sich vertieft mit der Komplexität im Gesamtzusammenhang auseinanderzusetzen.

## **4.2 Torspieler spezifische Daten und Advanced Metrics**

Die Datenerhebung der in Abschnitt 4.1 genannten Analysen erfolgte manuell durch die Autoren bzw. eine erweiterte Expertengruppe anhand individueller Definitionen für jede Aktion. Da eine individualisierte manuelle Erfassung im Rahmen des datenbasierten Scoutings nicht ökonomisch ist und auch nicht sinnvoll erscheint, muss auf die von professionellen Anbietern standardisiert erhobenen und zur Verfügung stehenden Daten zurückgegriffen werden. Im Rahmen dieser Arbeit erfolgte deshalb im ersten Schritt eine umfassende Marktanalyse der Anbieter mit besonderem Schwerpunkt auf Wyscout (vgl. Wyscout, 2023), Statsbomb (vgl. StatsBomb, 2023), Sporttec Solutions (vgl. bundesliga.com, 2023) und Stats Perform/Opta (vgl. Stats Perform, 2023), um festzustellen welche Daten angeboten werden. Zwischen den Anbietern gab es wie erwartet hinsichtlich der Datentiefe und der genauen Definitionen bzw. Benennung von Events große Unterschiede, auf die an dieser Stelle jedoch nicht im Detail eingegangen werden soll. Vielmehr ist es der Anspruch dieser Arbeit, ein vom Provider unabhängiges Modell zu entwickeln, mit dem die Leistungsfähigkeit eines Torspielers aus technisch-taktischer Sicht umfassend bewertet werden kann. Es ist logisch zu schlussfolgern, dass die Qualität der Ergebnisse maßgeblich von der Qualität, Quantität und Aussagekraft der verwendeten Daten abhängt. Durch ein providerunabhängiges Modell kann jedoch eine breite Anwendbarkeit in verschiedenen Ligen und Zielmärkten erreicht werden. Zudem besteht für Vereine, welche nicht über die finanziellen Mittel verfügen, um zusätzliche Daten zu erwerben, die Möglichkeit, die Kernideen des Modells zu nutzen, um aus den ihnen zur Verfügung stehenden Daten bessere Ergebnisse zu generieren.

Im Rahmen dieser Arbeit dienen die Kategorien der Torverteidigung, Raumverteidigung und Spieleröffnung, welche von Rechner/Memmert aus taktischer Perspektive abgeleitet und von Ziegler nahezu unverändert übernommen wurden (Offensivspiel anstelle von Spieleröffnung), als Orientierung zur Systematisierung der Daten. Jedoch erscheint die Terminologie des Mitspielverhaltens, die von Thaler und Krebs genutzt wird, angemessener. Daher wird im Rahmen dieser Arbeit zwischen den Kategorien der Zielverteidigung, dem Mitspielverhalten bei gegnerischem Ballbesitz und dem Mitspielverhalten im eigenen Ballbesitz unterschieden. Das Mitspielverhalten bei gegnerischem Ballbesitz entspricht im Wesentlichen der Raumverteidigung. Allerdings wird dieser Begriff verwendet, weil es für den Torspieler in dieser Spielphase nicht darum geht, den Raum zu verteidigen, sondern als Teil seiner Mannschaft aktiv am Ballgewinnspiel teilzunehmen, d.h. mitzuspielen um den Ball zu erobern. Das Mitspielverhalten im eigenen Ballbesitz entspricht weitgehend der Spieleröffnung bzw. dem Offensivspiel. Hier geht es jedoch nicht nur darum, den ersten Ball zu spielen, wie es der Begriff Spieleröffnung impliziert, sondern aktiv in das Angriffsspiel der eigenen Mannschaft eingebunden zu sein und wie ein elfter Feldspieler mitzuspielen. Unter Verwendung dieser grundlegenden Kategorisierung bietet Abbildung 6 eine Übersicht der derzeit auf dem Markt verfügbaren Torspielerspezifischen Daten und Advanced Metrics. Die englischen Begriffe der einzelnen Eventdaten wurden beibehalten, da diese Begriffe in der Regel auch von den Providern genutzt werden und unter Experten als Fachtermini geläufig sind.

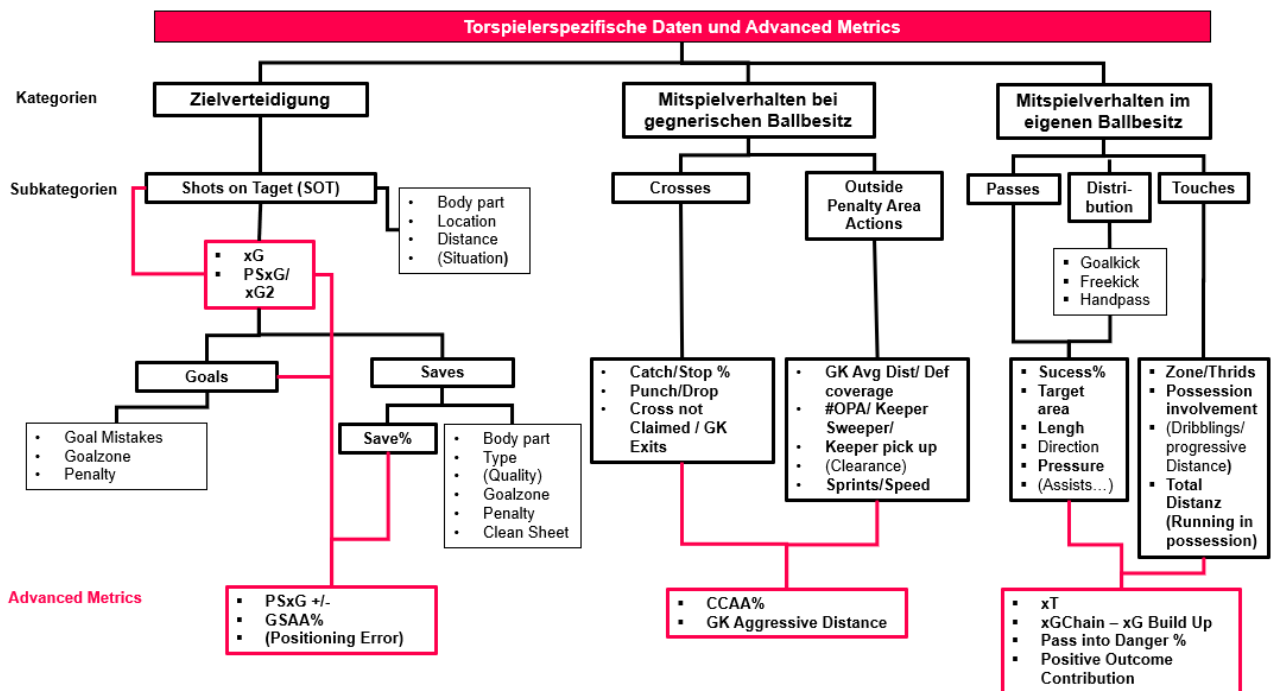


Abbildung 6: Torspielerspezifische Daten und Advanced Metrics Quelle: eigene Darstellung

In der Kategorie „Zielverteidigung“ bilden die "Shots on target" den Ausgangspunkt aller Torspieleraktionen. Diese werden je nach Anbieter bezüglich der Position des Schusses, der Distanz zum Tor, dem Körperteil (Kopfball/Schuss mit dem Fuß) oder sogar der Spielsituation aus der der Schuss entstanden ist, differenziert. Nachdem der Ball auf das Tor geschossen wurde, berechnen nahezu alle Anbieter den sogenannten Post-Shot xG, welcher neben den üblichen xG-Parametern auch die Qualität des Schusses berücksichtigt.

Im Ergebnis des Schusses wird logischerweise zwischen Tor und Gehalten (Saves) unterschieden. In der Subkategorie Tore wird zwischen Toren aus dem Spiel und Elfmetergegentoren unterschieden. Zum Teil wird auch die Zone innerhalb des Tores wo der Ball einschlägt angegeben oder die Anzahl der Fehler, die zu Gegentoren führen, aufgeführt. In der Subkategorie "Saves" ist die Fang-/Paradenquote das wesentliche Kriterium. Darüber hinaus bieten einige Anbieter detailliertere Informationen wie die Paradenart (1vs1, Hechten, Abtauchen, Ablenken, Fangen im Stand), die Qualität (festgehalten, nicht festgehalten) oder das Körperteil mit dem pariert wurde. Im weitesten Sinne fällt auch die Zu-Null-Quote (bzw. die Anzahl der Clean Sheets) in diese Subkategorie.

Bezüglich der Advanced Metrics für die Zielverteidigung hat sich noch keine wirklich auf dem Markt etabliert. Am aussagekräftigsten erscheint die Differenz aus Post-Shot xG und tatsächlich kassierten Gegentoren, auch als PSxG +/- bezeichnet, die von einigen Anbietern in ähnlicher Form zur Verfügung gestellt wird, aber auch selbst leicht berechnet werden kann. Darüber hinaus hat Statsbomb mit dem GSAA% - Goals Saved Above Average Percentage - Wert, der die Leistung des Torspielers im Vergleich zu den Erwartungen (berechnet aus  $(PSxG - \text{Tore}) / SOT$ ) berechnet, und dem Positioning Error-Wert, der angibt, wie weit der Torspieler bei einem Schuss im Durchschnitt von der optimalen Position entfernt ist, zwei weitere Metriken entwickelt. In Bezug auf letztgenannten ist anzumerken, dass es sehr lobenswert ist, dass sich endlich ein Anbieter mit dem Stellungsspiel, dem wichtigsten Erfolgsfaktor in der Zielverteidigung überhaupt, beschäftigt. Der Algorithmus und die Art und Weise, wie die vermeintlich optimale Position berechnet wird, erscheint jedoch noch nicht ausgereift, da wesentliche Elemente wie die Verkleinerung der Trefferfläche nicht berücksichtigt werden (vgl. Knutson, 2018).

In der Kategorie "Mitspielen bei gegnerischem Ballbesitz" werden Flanken und Aktionen außerhalb des eigenen Strafraums unterschieden. Im Hinblick auf Flanken ist die reine Anzahl der gefangenen Bälle in Prozent eine weit verbreitete Kennzahl. Einige Anbieter differenzieren zudem zwischen gefangenen, fallengelassenen und gefausteten Bällen sowie Versuchen, Flanken abzufangen, bei denen jedoch der Ball verfehlt wird. Eine Unterscheidung zwischen

Flanken, hohen Bällen und Querpässen oder ob diese nach Standards oder aus dem laufenden Spiel entstanden sind, ist in der Regel nicht möglich.

Die Aktionen außerhalb des eigenen Strafraums lassen sich durch die durchschnittliche Distanz der defensiven Aktionen oder den Bereich der defensiven Verantwortung, der durch die defensiven Aktionen eines Spielers während eines Spiels impliziert wird, bestimmen. Des Weiteren liefern als "Keeper-Sweeper" oder "#OPA" bezeichnete Ereignisse die reine Häufigkeit der Aktionen außerhalb des Strafraums. Ereignisse wie "Keeper Pick-Up", die die Aufnahme des Balls im Umschaltmoment zu eigenem Ballbesitz beschreiben, stehen in direktem Zusammenhang mit dem Mitspielverhalten bei gegnerischem Ballbesitz. Die Maximalgeschwindigkeit und die Anzahl der Sprints, lassen sich ebenfalls dieser Kategorie zuordnen, da diese Fähigkeiten insbesondere in Spielsituationen gefordert sind, bei denen der Torspieler mitspielen muss, um den Ball zu erobern oder zumindest zu klären. Gleiches gilt mit Abstrichen für die Laufleistung bei gegnerischem Ballbesitz, sofern diese separat verfügbar ist.

Statsbomb hat zwei spezifische Advanced Metrics für das Mitspielverhalten im gegnerischen Ballbesitz entwickelt, den "CCAA%-Wert" und die "GK Aggressive Distance". Der CCAA%-Wert versucht herauszufinden, wie aktiv die Torspieler beim Abfangen von hohen Bällen und Flanken in den Strafraum sind. Das "Claimables"-Modell definiert dabei zunächst die Wahrscheinlichkeit, dass ein Ball von und zu einer bestimmten Stelle erobert wird und bezieht dabei auch die Höhe der Bälle mit ein. Es bewertet dann die Torspieler auf der Grundlage ihrer Aktivität. Fleißige Torspieler, die sich von ihrer Linie lösen, um durch aktive Balleroberung eine Chance des Gegners zu verhindern, werden besser bewertet als solche, die ständig auf der Torlinie stehen. Die "GK Aggressive Distance" beschreibt die Präsenz und Aktivität des Torspielers sowie den defensiven Beitrag zur Mannschaftsleistung, indem die durchschnittliche Distanz von Clearances, Interceptions, Tackles und sonstigen Balleroberungen (ohne Saves und das Abfangen von Flanken) ermittelt wird (vgl. Knutson, 2018).

Innerhalb der Kategorie "Mitspielverhalten im eigenen Ballbesitz" können drei verschiedene Subkategorien unterschieden werden: Pässe, Spieleröffnungen und Ballkontakte. Die Spieleröffnungen beinhalten Abschlüsse, Abwürfe, das Ausrollen aus der Hand sowie die Eröffnungen nach Standardsituationen wie Abstöße oder Freistöße.

Pässe werden üblicherweise anhand ihrer Erfolgsquote in Prozent bewertet und hinsichtlich ihrer Länge (kurz, mittel, lang) sowie dem Zielraum (Drittel des Spielfelds) unterschieden. Vereinzelt werden auch Aspekte wie Progressivität, Richtung oder die Drucksituation aus der/in die gepasst wird, erhoben. Weitere Attribute wie Assists, expected Assists oder Key-Passes sind aus Torspieler-Sicht eher vernachlässigbar, da sie in der Praxis selten vorkommen. Spieleröffnungen werden teilweise nach dem gleichen Muster unterschieden.

Im Hinblick auf die Subkategorie "Ballkontakte" ist vor allem die Spielfeldzone, in der diese stattfinden, relevant. Opta liefert zudem mit "Sequences Involvement" einen interessanten Wert, der die Anzahl der Ballbesitzsequenzen beschreibt, an denen ein Spieler beteiligt war. In diesem Zusammenhang kann auch die progressive Distanz, die ein Torspieler mit einem Dribbling überbrückt hat, von Bedeutung sein. Ein weiterer Indikator für das Mitspielverhalten im eigenen Ballbesitz kann die Laufleistung im eigenen Ballbesitz sein, die sich im Wesentlichen aus der Aktivität im Freilaufverhalten ergibt. Es sollte jedoch beachtet werden, dass diese auch eine Doppelfunktion erfüllen kann, um bei einem möglichen Ballverlust in optimaler Position zu sein (z.B. Distanz zur Abwehrkette).

Im Bereich der Advanced Metrics für das Mitspielverhalten im eigenen Ballbesitz ist der xT-Wert ein weit verbreiteter Indikator, der die Bedeutung von einzelnen Spielaktionen im Gesamtkontext quantifiziert. Im Zusammenhang mit Torspielern wird dieser Wert bisher jedoch kaum berücksichtigt. Um den xT-Wert zu bestimmen, wird das Spielfeld in ein Raster aufgeteilt, und jeder Zelle wird eine Wahrscheinlichkeit zugeordnet, dass eine Aktion in dieser Zelle in den nächsten N Aktionen zu einem Tor führt. Der xT-Wert berücksichtigt nicht nur Zonen, in denen ein direktes Tor wahrscheinlicher ist, sondern auch solche, in denen ein Assist wahrscheinlicher ist. Aktionen wie Pässe oder Dribblings werden anhand ihres Start- und Endpunktes bewertet, indem der Unterschied im xT-Wert zwischen beiden Punkten berechnet wird. Ein negativer xT-Wert bedeutet, dass eine Aktion den Ball von einer gefährlicheren Zone wegbewegt hat. xGChain und xGBuild-Up sind weitere Metriken, die einen ganzheitlicheren Blick auf das Angriffsverhalten einer Mannschaft ermöglichen und den Beitrag eines einzelnen Spielers unabhängig von Toren und Assists bewerten (vgl. Lawrence, 2018). Torspielerspezifisch hat Statsbomb die Metriken "Pass into Danger%", welche angibt, wie oft der Pässempfänger unter Druck stand, und "Positive Outcome Contribution", welche angibt, wie häufig ein Spieler in Ballbesitzsequenzen involviert war, die mit einem positiven Ergebnis enden, entwickelt. Jedoch scheinen diese Metriken für die Bewertung der Fähigkeit des Mitspielverhaltens im Ballbesitz eher ungeeignet zu sein (vgl. Knutsen, 2018).

## **5 Methodik**

Das vorliegende Kapitel widmet sich der methodischen Vorgehensweise bei der Analyse von Daten im Anwendungsgebiet des Torspieler-Scoutings. Zunächst wird die Wahl des spezifischen Anwendungsgebiets erläutert und die konkreten Zielsetzungen des datenanalytischen Ansatzes definiert. Im Anschluss erfolgt eine detaillierte Beschreibung der verwendeten Datensätze. Im weiteren Verlauf des Kapitels werden die Berechnung torspielerspezifischer KPI's sowie die Erstellung von Clusteranalysen erörtert, die als Grundlage für den datenbasierten Scouting-Prozess dienen. Abschließend wird die grundlegende Vorgehensweise erläutert, um noch

individuellere und detailliertere Informationen aus den Daten extrahieren zu können, die bei der Entscheidungsfindung, welche Spieler auf die Short-List übernommen werden, von essentieller Bedeutung sind.

## **5.1 Wahl des Anwendungsgebietes & konkrete Zielstellung des datenanalytischen Ansatzes**

Das übergeordnete Ziel dieser Arbeit besteht darin, Lösungsansätze für ein datenbasiertes Torspieler-Scouting zu entwickeln, die das technisch-taktische Anforderungsprofil des Torspielers 4.0 in den Kernbereichen Zielverteidigung, Mitspielverhalten im eigenen Ballbesitz sowie Mitspielverhalten bei gegnerischem Ballbesitz (Raumverteidigung) erstmals ganzheitlich betrachten. Darüber hinaus soll das Modell unabhängig davon sein, von welchem Provider die genutzten Daten zur Verfügung gestellt werden. Somit kann das Modell in allen Ligen genutzt werden, in denen eine Spieldatenerhebung stattfindet. Unabhängigkeit bezieht sich hierbei auch auf die Technologie der Tracking-Datenerhebung sowie die detaillierten Definitionen und Benennung der Events. Dabei wird als selbsterklärend vorausgesetzt, dass die Qualität der Ergebnisse des Modells in direktem Zusammenhang mit der Qualität der Input-Daten steht. Das grundlegende Verständnis des Modells soll zudem Vereinen einen Low-Budget-Ansatz ermöglichen. Sofern für den jeweiligen Zielmarkt freizugängliche Daten zur Verfügung stehen, ist die Anwendung des Modells, abgesehen von Personalkosten für den Data Analyst und notwendigen Software-Kosten für ein BI-Tool (z.B. Tableau), ohne Zusatzkosten möglich. Ein solcher kostenneutraler Ansatz wird auch im Rahmen dieser Arbeit beispielhaft verwendet (vgl. 5.2). Falls dem Verein weitere Datensätze oder idealerweise API-Schnittstellen zu den Datenbanken führender Anbieter zur Verfügung stehen oder der Mehrwert des datenbasierten Scoutings erkannt wird und deshalb die Bereitschaft besteht zusätzliche Daten zuzukaufen, können diese selbstverständlich genutzt werden, was in der Regel zu einer Qualitätsverbesserung und einer größeren Detailtiefe führen sollte.

Das untergeordnete Ziel dieser Arbeit besteht darin, potenzielle Leser anzusprechen, deren primäres Interesse nicht das Scouting, sondern die (datenbasierte) Beschreibung der torspielerspezifischen Leistungsfähigkeit im Allgemeinen ist. Diese sollen dazu motiviert werden, den gegenwärtigen Status quo kritisch zu hinterfragen. Insbesondere sollen Medienvertreter dazu ermutigt werden, ihre aktuelle, tendenziell eindimensionale und häufig selbst im Bereich der Zielverteidigung unvollständige, Bewertung von Torspielerleistungen kritisch zu reflektieren. Zudem soll Datenprovidern ein besseres Verständnis vermittelt werden, welche KPIs, die ohnehin für Feldspieler erhoben werden, auch für Torspieler von zentraler Bedeutung sind und durch die Kombination welcher torspielerspezifischen Daten neue aussagekräftige Advanced Metrics entwickelt werden können. Außerdem sollen weitere torspielerspezifische Leistungsfaktoren



identifiziert werden, die aus Sicht des Autors mithilfe aktueller Technologien problemlos standardisiert erfasst werden könnten und einen erheblichen Mehrwert in der Analyse der Torspielerleistung bieten würden.

## 5.2 Beschreibung der Datensätze

Im Zuge der Modellentwicklung innerhalb dieser Arbeit wurden primär Eventdaten von der öffentlich zugänglichen Online-Plattform fbref.com herangezogen. Diese Plattform stellt in Zusammenarbeit mit Opta "fortgeschrittene analytische Daten" für mehr als 20 Wettbewerbe zur Verfügung (vgl. fbref.com, 2023). In Tabelle 1 sind die aus Sicht des Autors torspielerrelevanten Daten, die auf dieser Plattform zugänglich sind, den jeweiligen Kategorien (vgl.4.2) zugeordnet. Die Oberbegriffe der Events sind dabei fett gedruckt, während die jeweiligen Attribute normal geschrieben sind, um eine bessere Übersichtlichkeit zu gewährleisten. Neben den bekannten Kategorien, Zielverteidigung, Mitspielverhalten bei gegnerischem Ballbesitz und Mitspielverhalten im eigenen Ballbesitz, wurden in einer weiteren Spalte weitere für das Scouting relevante Daten wie Stammdaten, Spielzeit und Teamerfolg aufgeführt.

Tabelle 1: Torspieler relevante Eventdaten fbref.com Quelle: eigene Darstellung

<b>Stammdaten, Einsatzzeit &amp; Teamerfolg</b>	<b>Zielverteidigung</b>	<b>Mitspielverhalten bei gegnerischem Ballbesitz</b>	<b>Mitspielverhalten im eigenen Ballbesitz</b>
Player	<b>Goals against</b> / 90 (FK, CK, OG)	<b>Crosses</b> /90 (Opp, Stp, Stp%)	<b>Launched</b> /90 (Cmp, Att, Cmp%)
Nation	<b>SoTA</b> / 90	<b>Sweeper</b> (#OPA/90, AvgDist)	<b>GK Passes</b> /90 (Att, Thr, Launch%, AvgLen)
Position	<b>Saves</b> / 90 /%	Interceptions/90	<b>Goal Kicks</b> /90 (Att, Launch%, AvgLen)
Current age/ Year of birth	<b>Clean Sheets</b> / %	Clearances/90	<b>Passing Total</b> /90 (Cmp, Att, Cmp%, TotDist, ProgDist, KP, into1/3)
<b>Playing time</b> (MP, Starts, Min, 90s)	<b>Penalty Kicks</b> (PKatt, PKA, PKsv, PKm, Save%)	Tackles/90 (Won)	<b>Passing Short/Medium/Long</b> /90 (Cmp, Att, Cmp%)
Wins/Draws/Losses	<b>PSxG</b>		<b>Pass Types</b> /90 (Live, Dead, FK, Sw)
Team Success/90 (PPM, onG, onGA,	<b>PSxG/SoT</b>		Shot/Goal Creation Actions/90

+/- (onG-onGA)/90, On-Off)			
TeamSucces (xG) (onxG,onxGA, onxG+/-/90,xG On-Off	<b>PSxG +/- /90</b>		<b>Touches/90</b> (Def Pen, Def 3 <sup>rd</sup> , Mid 3 <sup>rd</sup> , Att 3 <sup>rd</sup> , Live)
			<b>Carries/90</b> (TotDist, PrgDist, Mis, Dis)
			<b>Passes Received/ 90</b>

Im Rahmen dieser Arbeit werden exemplarisch die Daten der 1. Bundesliga (Saison 2022/2023) verwendet. Zum Zeitpunkt der Datenerhebung waren 24 Spieltage, also insgesamt 216 Spiele, absolviert. In diesem Zeitraum kamen 33 verschiedene Torspieler zum Einsatz, wobei Yann Sommer aufgrund seines Wintertransfers von Borussia Mönchengladbach zum FC Bayern München als einziger Torspieler für zwei verschiedene Mannschaften aktiv war. Aus Gründen der Validität werden diese Einsätze getrennt voneinander betrachtet.

Zur Ergänzung dieses Datensatzes wurden Trackingdaten wie Sprints/90, intensive Läufe/90, Laufdistanz/90 und Geschwindigkeit, welche auf der offiziellen Seite der Bundesliga (vgl. bundesliga.com, 2023a) frei zugänglich sind, sowie Transfermarkt-Daten wie Vertragslaufzeit und Marktwert von transfermarkt.de (vgl. transfermarkt.de, 2023) hinzugefügt.

Um die Vergleichbarkeit der Leistungen der Torspieler zu gewährleisten, wurden bei allen quantitativen, volumenbasierten Metriken die Werte pro 90 Minuten (p90) verwendet. Dies stellt einen guten Ausgangspunkt für einen fairen Vergleich dar, allerdings bleibt die Mannschaftsleistung, die einen großen Einfluss auf die reine Anzahl der Aktionen hat, bei diesem Verfahren unberücksichtigt. Grundsätzlich wird empfohlen, die Werte auf Spielerebene durch einen oder mehrere Werte auf Mannschaftsebene zu bereinigen, um eine "neutrale" Basis zu schaffen und den Einfluss der Mannschaftsleistung zu eliminieren. Zum Beispiel können die Anzahl der gespielten Pässe auf ein ballbesitzneutrales Niveau (50% Ballbesitz der Mannschaft) angepasst, Fangquoten pro X zugelassene Schüsse/Flanken ermittelt oder diese ins Verhältnis zur Qualität der Abschlüsse gesetzt werden. Leider stehen auf fbref.com keine konkreten Teamdaten zur Verfügung, die sich nur auf den Zeitraum beziehen in denen ein bestimmter Spieler auf dem Platz war, weshalb im Rahmen dieser Arbeit auf diese tiefere Datenbereinigungen verzichtet wurde. Es liegt auf der Hand, dass bei Verwendung der p90-Werte Spieler mit geringer Einsatzzeit zu extremen Werten neigen können. Jedoch können diese Spieler

mithilfe von Filterfunktionen während jeder Phase des Scouting-Prozesses ausgeschlossen oder bei Bedarf tiefer analysiert werden.

### **5.3 Berechnung von KPI's und Clusteranalyse zur Kategorisierung von Torspielern**

Um einen ersten Überblick zur Leistungsfähigkeit der Torspieler in den Kategorien Zielverteidigung, Mitspielverhalten bei gegnerischem Ballbesitz und Mitspielverhalten im eigenen Ballbesitz zu erhalten, wurde ein übergreifender KPI aus den Rohdaten der jeweiligen Kategorie gebildet. Die Gewichtung der einzelnen Elemente/Aktionen/Metriken kann individuell anhand der eigenen Spielidee und den vereinspezifischen Anforderungen angepasst werden, um präferierte Verhaltensweisen und Bedürfnisse zu berücksichtigen.

Im Rahmen dieser Arbeit wird der KPI "Zielverteidigung" hauptsächlich aus den advanced Metrics PSxG+/-, PSxG/SoT und der Fangquote (Save%) gebildet. Der KPI "Mitspielverhalten bei gegnerischem Ballbesitz" setzt sich aus #OPA, Avg Dist, Tkl, Int., Clr, Crosses stopped, Stop%Crosses, Sprints, Intensiven Läufen und der Geschwindigkeit zusammen. Der KPI "Mitspielverhalten im eigenen Ballbesitz" beinhaltet Live Touches & Passes, Passquote, Touches im mittleren und ersten Drittel, progressive Distanz und Gesamt-Laufdistanz. Die Gewichtung der einzelnen Kategorien wird aus dem Anforderungsprofil des Torspielers 4.0 abgeleitet und erhebt keinen Anspruch auf allgemeine Gültigkeit.

Im nächsten Schritt des Torspieler-Scoutingprozesses wurde eine mehrstufige Clusteranalyse unter Verwendung von Tableau durchgeführt, um Ähnlichkeiten, Unterschiede und Muster zwischen den spielspezifischen Werten der einzelnen Spieler, die als Indikator für die Leistungsfähigkeit betrachtet werden, zu erkennen. Die Clusteranalyse ist eine Methode des unsupervised Data Mining, die darauf abzielt, Objekte in homogene Gruppen/Cluster zu teilen, wobei ähnliche Objekte denselben Clustern zugeordnet werden. Obwohl es verschiedene Methoden der Clusteranalyse gibt, wird es zum Verständnis dieser Arbeit als ausreichend erachtet bei dieser grundlegenden Definition zu bleiben (vgl. Maiwald, 2013, S.9).

Die erste Stufe der Clusteranalyse erfüllt eine Doppelfunktion. Erstens ermöglicht sie die unkomplizierte Identifikation von Spielern mit ähnlichen Eigenschaften. Somit können potenzielle Neuzugänge schnell ausfindig gemacht werden, wenn die derzeitige Nummer eins den Verein verlässt oder ersetzt werden soll. Da die Ähnlichkeit gewährleistet ist, kann erwartet werden, dass diese potenziellen Neuzugänge einfacher in das Team integriert werden können, ohne dass die grundlegende Spielidee in den verschiedenen Spielphasen verändert werden muss. Zudem können Spieler identifiziert werden, die ein ähnliches Profil haben, aber in ihrem aktuellen Umfeld in bestimmten Bereichen signifikant besser sind. Zweitens lässt sich aufgrund des Wissens über

die Komplexität des Anforderungsprofils des Torspielers sowie dessen historischer Entwicklung vermuten, dass ein Zusammenhang zwischen dem Spielerprofil (d.h. der Clusterzugehörigkeit) und dem Teamerfolg besteht. Konkret bedeutet dies, dass der Torspielertyp einen Einfluss auf die erzielten Punkte, die Tordifferenz und/oder die Differenz der erwarteten Tore hat. Um diese Hypothese zu überprüfen, wird eine Regressionsanalyse durchgeführt. Sollte sich die Hypothese bestätigen, könnte dies bedeuten, dass Torspieler aus einem bestimmten Cluster gegenüber anderen bevorzugt werden sollten, sofern die Fähigkeiten, die diese Spieler auszeichnen, im Rahmen der eigenen Spielphilosophie gewinnbringend eingesetzt werden können. Bei der Untersuchung des Zusammenhangs zwischen der Leistung der Mannschaft und den spielspezifischen Werten des Torspielers ist einschränkend die Frage der Kausalität zu berücksichtigen. Es stellt sich die Frage, ob die Mannschaft erfolgreich ist, weil die spielspezifischen Werte des Torspielers gut sind oder ob die spielspezifischen Werte des Torspielers gut sind, weil die Mannschaft erfolgreich ist. Eine abschließende Beantwortung dieser Frage gestaltet sich jedoch als äußerst schwierig, da es in der Praxis viele Abhängigkeiten zwischen diesen Variablen gibt.

Im dritten Schritt des datenbasierten Torspielerscoutings wird eine ergänzende Methode des Clusteringverfahrens verwendet, die darin besteht, Torspieler anhand einer leistungsabhängigen Kombination der bekannten Kategorien und den zugehörigen KPIs zu charakterisieren. Dadurch kann ein Torspieler, der einem bestimmten Cluster zugewiesen wurde, durch eine gewichtete Kombination von KPIs repräsentiert werden. Zum Beispiel kann ein Torspieler, der als "sehr guter Torspieler mit Stärken in der Zielverteidigung und im Mitspielverhalten im eigenen Ballbesitz" eingestuft wurde, zu 50% einen mitspielenden Torspieler, zu 30% einen Linientorwart (Zielverteidigung) und zu 20% einen Sweeper (Mitspielverhalten bei gegnerischem Ballbesitz) repräsentieren. Damit diese Methode angewendet werden kann, müssen alle KPI-Werte im gleichen Wertebereich abgebildet werden. In dieser Arbeit wurden die KPI-Werte auf den Bereich zwischen 0 und 100 normiert, basierend auf dem höchsten Wert. Durch diese detaillierte Analyse der Spielweise können die Wichtigkeit und Konformität eines Torspielers mit der Spielidee genauer quantifiziert werden.

Im vierten Schritt der Analyse werden die Rohdaten der Events einer detaillierteren Interpretation unterzogen, um präzisere Informationen über spezifische Fähigkeiten zu erhalten, die für die eigene Spielidee oder die individuellen/vereinspezifischen Anforderungen im Spiel von besonderer Bedeutung sind. Denn das Ziel des Scouting ist nicht, den bestmöglichen Spieler zu finden, sondern den Spieler, der aufgrund seiner Stärken am besten zum jeweiligen Verein passt.

Im Rahmen dieser Arbeit wurden hauptsächlich mit Tableau erstellte Streudiagramme verwendet, die zwei Metriken/Events auf der x- und y-Achse kombinieren. Dabei werden überwiegend

volumenbasierte und qualitative Werte kombiniert, wobei in der Regel für jede Metrik eine durchgezogene Median-Linie, Quartile und/oder eine Trendlinie, die Korrelationen verdeutlicht, eingefügt wird. Dadurch kann man visuell auf den ersten Blick feststellen, ob ein Spieler in diesem Bereich über-/unterdurchschnittlich ist oder sogar zu den besten Spielern gehört, die in der Regel "oben rechts" zu finden sind. In Ausnahmefällen wird durch die Form oder Farbe eine weitere Metrik ergänzt, sofern diese aus Sicht des Autors einen Mehrwert in der Interpretation der Daten liefert. Eine weitere wichtige Visualisierungsform in diesem Schritt sind Balkendiagramme, mit denen sich reine Häufigkeiten oder Attribute von einzelnen Events übersichtlicher darstellen lassen.

Im fünften und letzten Schritt des datenbasierten Torspieler-Scoutings wird ein Dashboard erstellt, auf dem alle relevanten Daten eines Spielers dargestellt werden. In der Praxis sind hierbei hauptsächlich zwei Schritte zu beachten: Zum einen müssen Filterfunktionen implementiert werden, mit denen die potenziellen Spieler anhand bestimmter Kriterien wie beispielsweise Alter, Vertragslaufzeit, Marktwert, Spielzeit, Spielercluster oder Leistungsfähigkeit in einem bestimmten Teilbereich begrenzt werden können oder ein Vergleich zwischen wenigen Spielern gezogen werden kann. Zum anderen muss der ausgewählte Spieler durch "Dashboard-Aktionen" hervorgehoben werden, so dass er auf allen Teilbereichen des Dashboards sofort erkennbar ist. Durch die Verwendung des Dashboards können somit interessante Spieler identifiziert und eine erste Short-List potenzieller Kandidaten erstellt werden. Diese Liste kann anschließend durch qualitative Video- oder Livescouting weiter reduziert werden.

## **6 Ergebnisse und Diskussion**

Im Folgenden werden die Ergebnisse des in 5.3 beschriebenen Scoutingprozesses mit Hilfe von Tableau dargestellt, erläutert bzw. interpretiert und diskutiert.

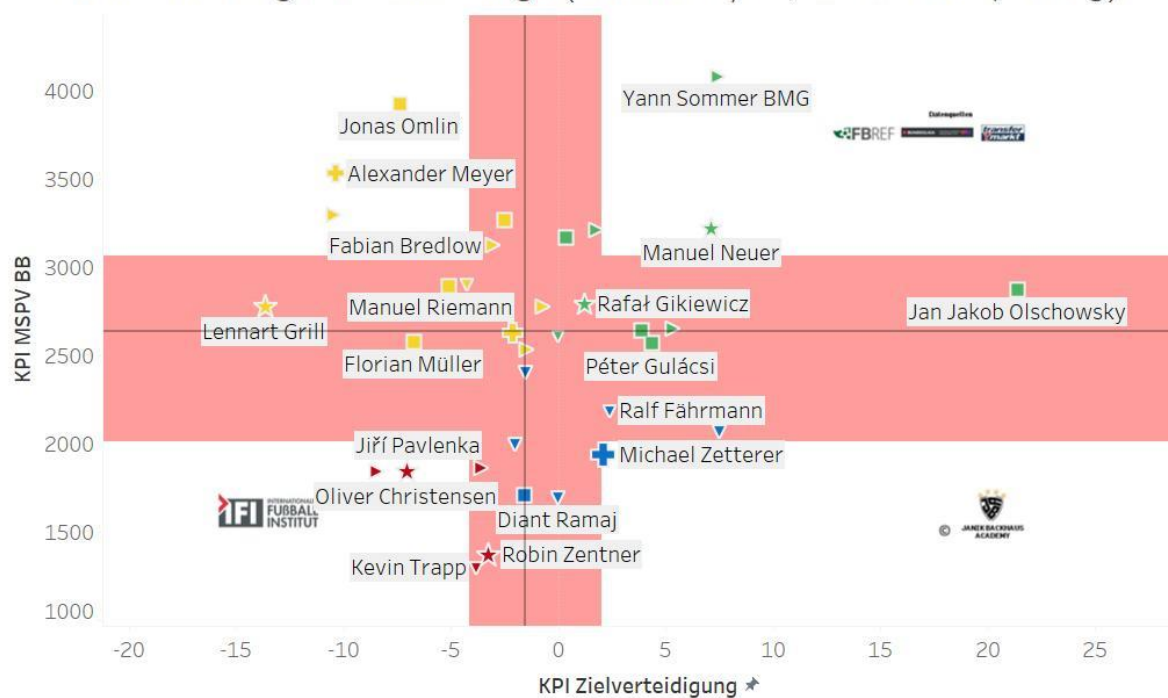
### **6.1 Darstellung mit Hilfe von Tableau und Erläuterung der gewonnenen Ergebnisse**

#### **6.1.1 Clusteranalyse nach Leistungsfähigkeit**

Wie bereits in Abschnitt 5.3 erläutert wurde, bildet die Clusteranalyse den Ausgangspunkt des Scouting-Prozesses. Abbildung 7 zeigt die Ergebnisse dieser Analyse. Dabei wurde der KPI-Zielverteidigung auf der X-Achse abgetragen, wobei eine Position weiter rechts eine höhere Leistungsfähigkeit bedeutet. Die Y-Achse hingegen repräsentiert den KPI-Mitspielverhalten im eigenen Ballbesitz, wobei eine höhere Position auf der Y-Achse eine bessere Bewertung des KPIs anzeigt. Die Form der Symbole beschreibt den KPI-Mitspielverhalten bei gegnerischem Ballbesitz in einer hierarchischen Ordnung (+ > ☆ > ■ > ▶ > ▼). Die Farben der Symbole

repräsentieren die verschiedenen Cluster, denen die jeweiligen Torspieler zugeordnet wurden.

### Clusterbildung TSP Bundesliga (Saison 22/23, Stand 24. Spieltag)



#### Player (Cluster)

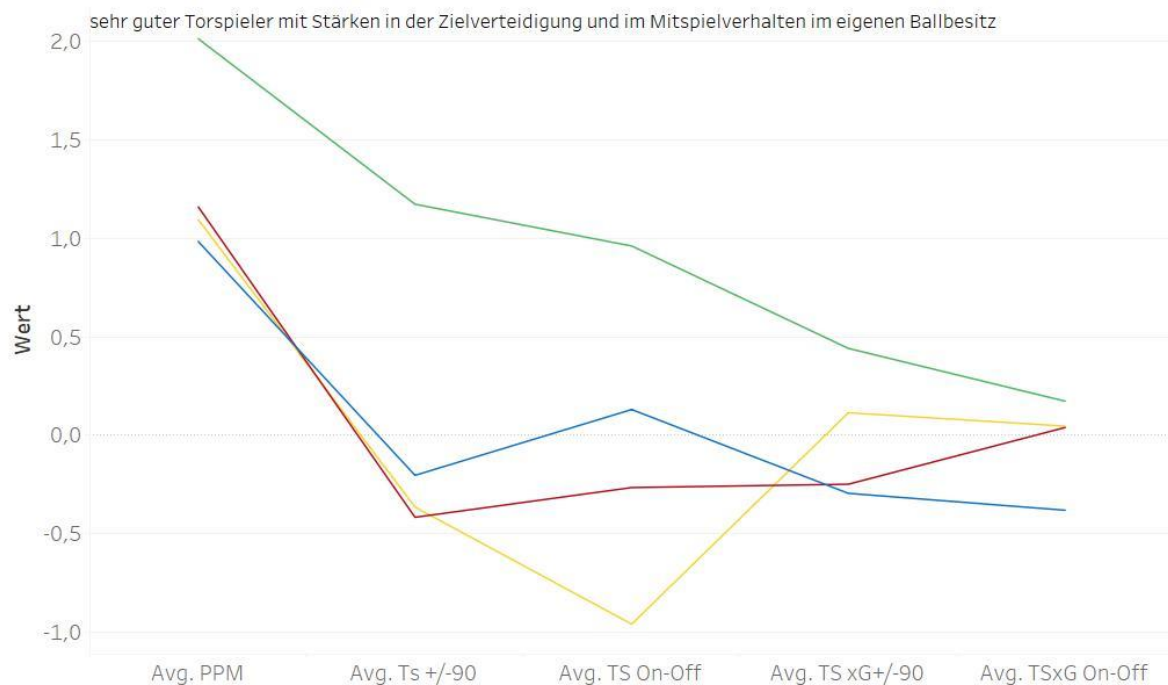
- durchschnittlicher Torspieler mit Stärken in der Zielverteidigung aber Defiziten im Mitspielverhalten bei eigenem Ballbesitz
- unterdurchschnittlicher Torspieler mit Defiziten in der Zielverteidigung und im Mitspielen bei eigenem Ballbesitz
- sehr guter Torspieler mit Stärken in der Zielverteidigung und im Mitspielverhalten im eigenen Ballbesitz
- durchschnittlicher Torspieler mit Stärken im Mitspielverhalten im eigenen Ballbesitz aber Defiziten in der Zielverteidigung

Abbildung 7: Player Cluster Quelle: eigene Darstellung

Insgesamt wurden die Torspieler aufgrund ihrer Leistung vier verschiedenen Clustern zugeordnet. Cluster 1 (grün, 10x) beinhaltet sehr gute Torspieler mit Stärken sowohl in der Zielverteidigung als auch im Mitspielverhalten im eigenen Ballbesitz. Cluster 2 (blau, 7x) beinhaltet durchschnittliche Torspieler mit Stärken in der Zielverteidigung, jedoch Defiziten im Mitspielverhalten bei eigenem Ballbesitz. Cluster 3 (gelb, 12x) umfasst durchschnittliche Torspieler mit Stärken im Mitspielverhalten im eigenen Ballbesitz, aber Defiziten in der Zielverteidigung. Cluster 4 (rot, 5x) beinhaltet unterdurchschnittliche Torspieler mit Defiziten sowohl in der Zielverteidigung als auch im Mitspielverhalten im eigenen Ballbesitz. Es ist anzumerken, dass die Leistungsfähigkeit im Mitspielverhalten bei gegnerischem Ballbesitz insbesondere innerhalb der Cluster 1, 3 und 4 divergiert. Im Falle des Cluster 2 kann abgeleitet werden, dass diese Spieler auch im Mitspielverhalten bei gegnerischem Ballbesitz schwach sind, was typisch für klassische Linientorhüter ist. Es ist jedoch zu erwähnen, dass die Ausnahmen Diant Ramaj und Michael Zetterer aufgrund ihrer geringen Einsatzzeit gesondert betrachtet werden sollten.

In der vorliegenden Arbeit wurde anschließend die Hypothese untersucht, ob die Clusterzugehörigkeit eines Torspielers einen signifikanten Einfluss auf den Erfolg des Teams hat, um die Validität der selbstberechneten KPI's zu überprüfen. Hierzu wurde eine Regressionsanalyse durchgeführt, in der die Clusterzugehörigkeit als abhängige Variable und verschiedene Teamerfolgsvariablen (Punkte pro Spiel (PPM), Tordifferenz (TS +/-90), Nettotore pro 90 Minuten der Mannschaft während der Spieler auf dem Platz war abzüglich der Nettogegentore pro 90 Minuten während der Spieler nicht auf dem Platz stand (TS On-Off) und die jeweiligen xG-Werte) als unabhängige Variablen betrachtet wurden. Die Ergebnisse dieser Analyse ist in den Abbildung 8 dargestellt.

### Regressionsanalyse (Cluster vs. Team Erfolg) grafische Darstellung



#### Player (Cluster)

- durchschnittlicher Torspieler mit Stärken in der Zielverteidigung aber Defiziten im Mitspielverhalten bei eigenem Ballbesitz
- unterdurchschnittlicher Torspieler mit Defiziten in der Zielverteidigung und im Mitspielen bei eigenem Ballbesitz
- sehr guter Torspieler mit Stärken in der Zielverteidigung und im Mitspielverhalten im eigenen Ballbesitz
- durchschnittlicher Torspieler mit Stärken im Mitspielverhalten im eigenen Ballbesitz aber Defiziten in der Zielverteidigung

**Abbildung 8:** Regressionsanalyse grafische Darstellung Quelle: eigene Darstellung

Die Ergebnisse der Regressionsanalyse zeigen einen signifikanten Unterschied zwischen Spielern, die dem Cluster 1 (Stärken im Mitspielverhalten im eigenen Ballbesitz und in der Zielverteidigung) zugeordnet wurden, und allen anderen Clustern. Im Durchschnitt gewinnen diese Spieler mit ihren Mannschaften 0,85 - 1,03 mehr Punkte pro Spiel und die Tordifferenz pro 90 Minuten ist um 1,37 - 1,59 Tore besser. Die Differenzen zwischen den anderen Clustern sind hingegen nicht signifikant. Daraus lässt sich ableiten, dass es aus rein erfolgsorientierter Sicht keinen nennenswerten Unterschied macht, ob ein Torspieler mit Stärken in der Zielverteidigung

und Defiziten im Mitspielverhalten (Cluster 2) oder umgekehrt (Cluster 3) verpflichtet wird. Entscheidend ist in diesem Fall das Entwicklungspotenzial im Defizit-Bereich, um in Cluster 1 zu gelangen. Es ist jedoch zu beachten, dass das Defizit in den Werten auch durch die Stärke/Schwäche bzw. die Spielidee seiner aktuellen Mannschaft bedingt sein kann.

### 6.1.2 Clusteranalyse nach Torspielertyp

Die in Abschnitt 5.3 beschriebene ergänzende Methode der Clusteranalyse, bei der Torspieler durch eine leistungsfähigkeitsabhängige, gewichtete Kombination der Kategorien/KPI's repräsentiert werden, ist von theoretischer Natur. Diese Darstellungsform ermöglicht jedoch eine detailliertere Beschreibung des exakten Spielerprofils des jeweiligen Torspielers. In diesem Schritt geht es deshalb nicht darum, die Qualität eines Spielers zu beurteilen, sondern seine Spielweise genauer zu beschreiben. Die Spielertypenanalyse ist in Abbildung 9 dargestellt.

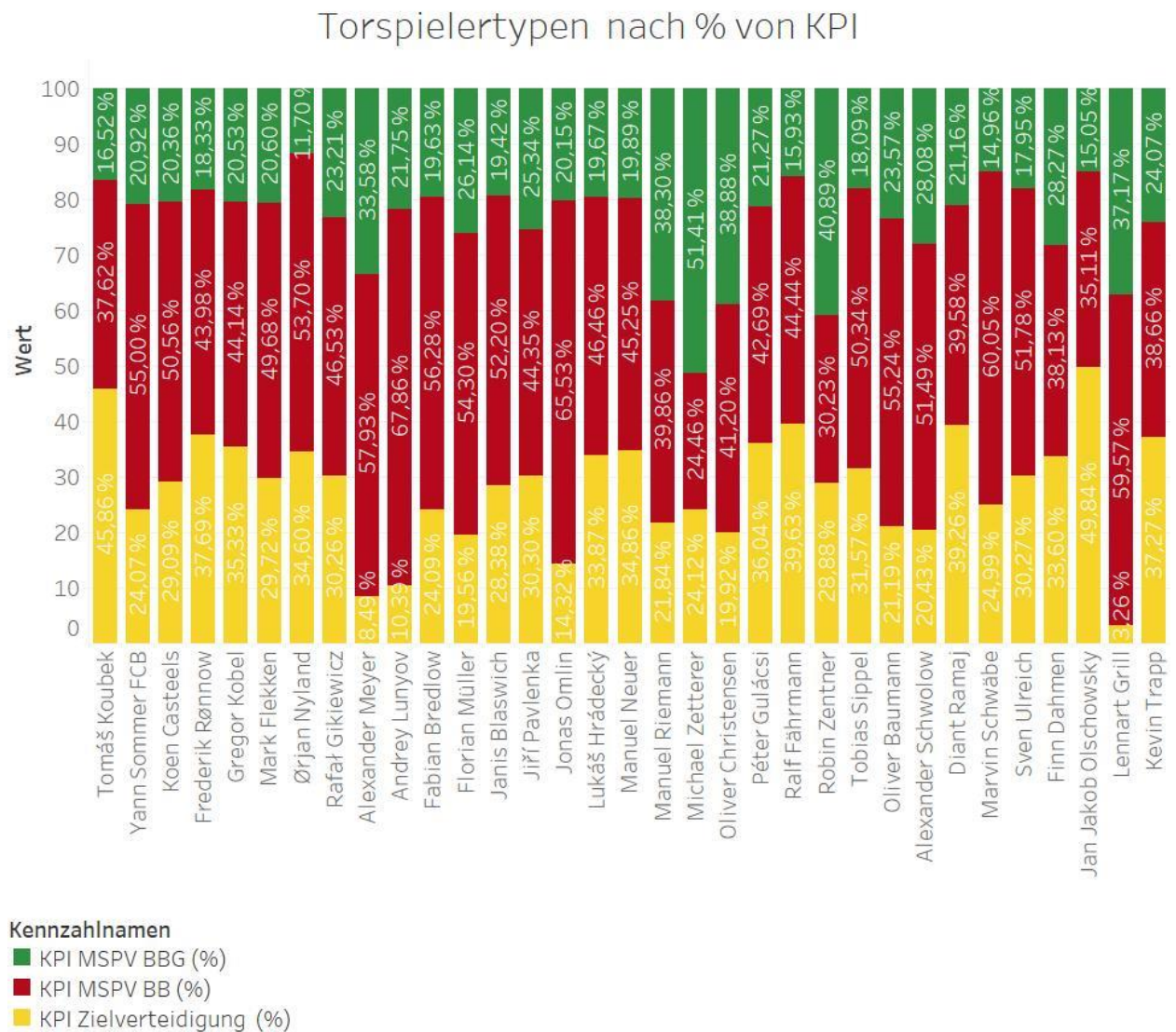


Abbildung 9: Torspielertypen nach % von KPI Quelle: Eigene Darstellung



Eine erste Analyse der Abbildung zeigt, dass die meisten Torspieler überwiegend mitspielende Torspieler sind. Die Zielverteidigung folgt zumeist und das Mitspielverhalten bei gegnerischem Ballbesitz nimmt den geringsten Anteil ihres Spielerprofils ein. Es ist interessant, nicht nur die detaillierte Gewichtung der einzelnen Faktoren zu betrachten, sondern vor allem die Ausnahmen von der Regel zu erkennen. So ist beispielsweise erkennbar, dass die Zielverteidigung im Spielerprofil von Jonas Omlin und Oliver Baumann nur einen sehr geringen Anteil einnimmt. Diese Spieler wären für Mannschaften, die viele gegnerische Torschüsse zulassen und wenig eigenen Ballbesitz haben, vermutlich eher ungeeignet. Gleiches gilt für Manuel Riemann und Oliver Christensen, die aufgrund der nahezu ausgeglichenen Gewichtung von offensivem und defensivem Mitspielverhalten aber eher für Mannschaften mit weniger eigenem Ballbesitz, die jedoch nicht tief verteidigen, sondern bewusst Sweeper-Keeper-Aktionen provozieren, geeignet wären. Nach diesem Muster könnte man jeden einzelnen Spieler mit dem eigenen Anforderungsprofil vergleichen und so die potenziellen Neuzugänge weiter eingrenzen. Ein theoretisches Idealprofil innerhalb dieses Darstellungsmodells wäre das des kompletten Torspielers mit einer Gleichverteilung von 33,3 % pro Kategorie. Kevin Trapp, Finn Dahmen, Manuel Neuer, Gregor Kobel, Lukas Hradecky und Diant Ramaj kommen diesem Ideal am nächsten. Es ist jedoch zu beachten, dass Gleichverteilung in diesem Kontext nicht zwangsläufig gleich gut bedeutet, sondern auch gleich schlecht bedeuten kann. Dies liegt daran, dass es sich lediglich um eine Beschreibung des Spielerprofils handelt. Dies unterstreicht die Notwendigkeit mehrerer Schritte im datenbasierten Scoutingprozess.

### 6.1.3 Detailanalyse Zielverteidigung

Um eine umfassendere Analyse der Fähigkeiten in der Zielverteidigung durchzuführen, ist es sinnvoll, den KPI "Zielverteidigung" als Ausgangspunkt zu betrachten. In Abbildung 10 werden die Torspieler, die mindestens 20% der Spielzeit absolviert haben, in absteigender Reihenfolge aufgeführt. Diese Filterfunktion wird in allen nachfolgenden grafischen Darstellungen angewendet, um eine höhere Übersichtlichkeit und Validität zu gewährleisten.

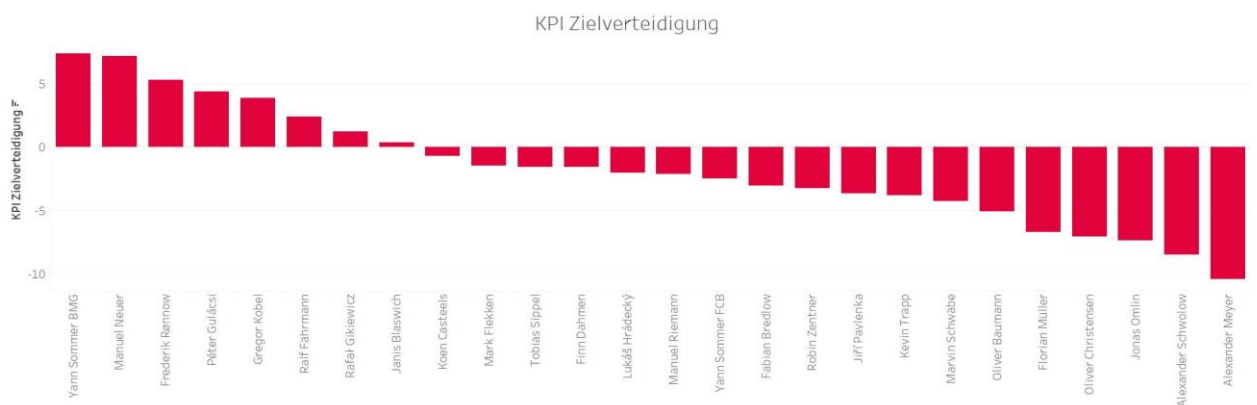


Abbildung 10: KPI Zielverteidigung Quelle: eigene Darstellung

Als zweiten Schritt empfiehlt sich die Fangquote in Bezug zur Anzahl der Schüsse auf das Tor zu setzen, wie in Abbildung 11 dargestellt.

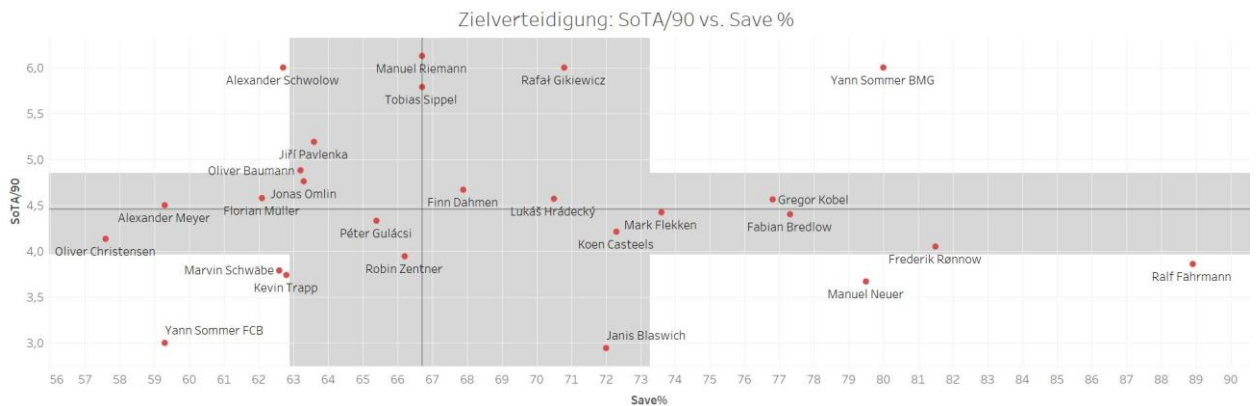


Abbildung 11: SoTA/90 vs. Save% Quelle: eigene Darstellung

Hinsichtlich des Scouting von Spielern ist es interessant, jene Spieler zu betrachten, die sich im rechten oberen Bereich befinden und trotz einer hohen Anzahl an Schüssen auf das Tor eine hohe Fangquote aufweisen. Aus Sicht der Mannschaft ist es jedoch vorteilhafter, Spieler im rechten unteren Bereich zu haben, d.h. wenige Torschüsse des Gegners zuzulassen, jedoch einen Torspieler mit einer guten Save-Quote zu haben, um diese zu parieren. Dementsprechend scheint eine Position im linken unteren Bereich ein Indikator für eine schwache Leistungsfähigkeit zu sein. Es ist jedoch zu beachten, dass dies auch mit der Qualität der zugelassenen Chancen zusammenhängen kann. Daher ist eine weitere Analyse notwendig. Die Abhängigkeit der Leistung des Torspielers von der Mannschaftsleistung wird besonders deutlich, wenn man die Performance von Yann Sommer in seinen beiden Vereinen betrachtet. Während er in Gladbach mit vielen Schüssen konfrontiert war und eine hohe Paradenquote aufwies, war es bei Bayern München das umgekehrte Extrem. Um die Qualität der zugelassenen Torschüsse zu überprüfen, wird als nächster Schritt der Post Shot xG-Wert pro 90 Minuten durch die Anzahl der Schüsse auf das Tor pro 90 Minuten geteilt. Ein höherer Wert ist ein Indikator für gefährlichere Torschüsse bzw. größere Chancen des Gegners. Dieser Wert wird mit der wohl aussagekräftigsten Advanced Metric, dem PSxG +/- /90, ins Verhältnis gesetzt, der die Differenz zwischen erwarteten und tatsächlichen Gegentoren beschreibt (vgl. Abb. 12).

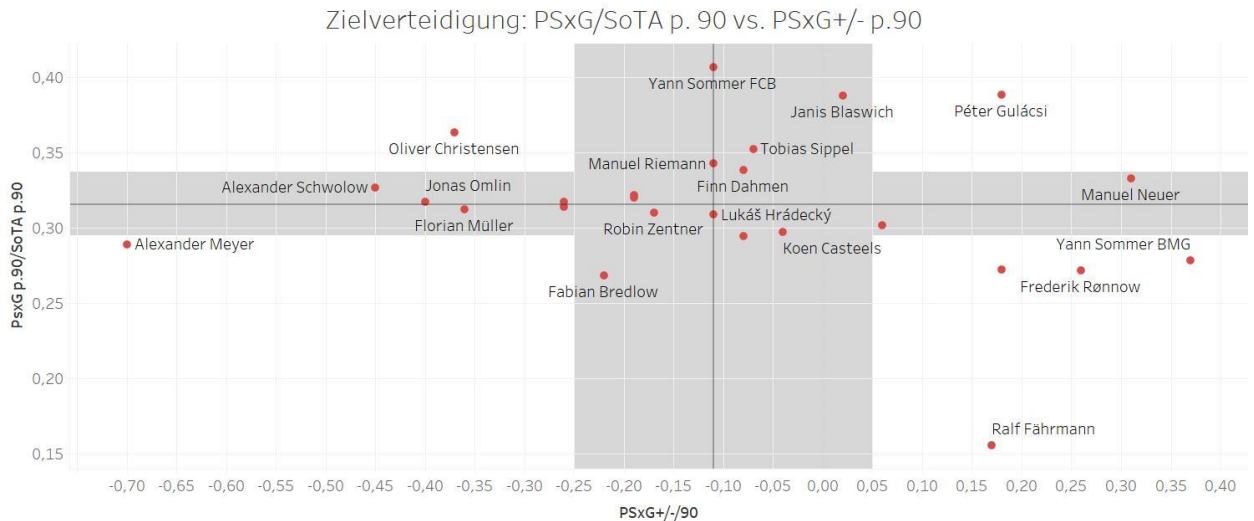


Abbildung 12: PSxG/SoTA p.90 vs. PSxG+/- p.90 Quelle: eigene Darstellung

Die hier präsentierte rein qualitative Analyse dient dazu, die Ergebnisse von Abbildung 11 in einen angemessenen Kontext zu stellen. Bei der Betrachtung von Yann Sommers Leistung wird deutlich, dass die von ihm zugelassenen Torschüsse während seiner Zeit bei Bayern die gefährlichsten waren und dass sein PSxG+/-/90-Wert genau im Median liegt, was bedeutet, dass er nur geringfügig mehr Gegentore zugelassen hat, als erwartet wurde. Während er bei Gladbach sogar den besten PSxG+/-/90-Wert aufwies, waren die zugelassenen Torschüsse eher ungefährlich. Bei der Interpretation gelten dieselben Richtlinien wie bei der vorherigen Abbildung: Eine Position rechts oben ist ein Indikator für eine außergewöhnlich gute Leistungsfähigkeit, während links unten eher auf Defizite hindeutet. Sollte sich ein Spieler der eigenen Mannschaft in diesem Quadranten befinden, besteht zweifellos Handlungsbedarf und es ist angebracht, Alternativen in Betracht zu ziehen. Um den Faktor der Konstanz noch stärker zu betonen, kann man optional die nicht pro 90 Minuten adjustierten Werte heranziehen. Diese Vorgehensweise birgt jedoch die Gefahr, dass konstant gute Torspieler mit viel Spielzeit tendenziell begünstigt werden, während konstant schwache Torspieler mit viel Spielzeit und Torspieler mit vergleichsweise wenig Spielzeit, die in der Regel zur Mitte tendieren, benachteiligt werden (vgl. Abb. 13).

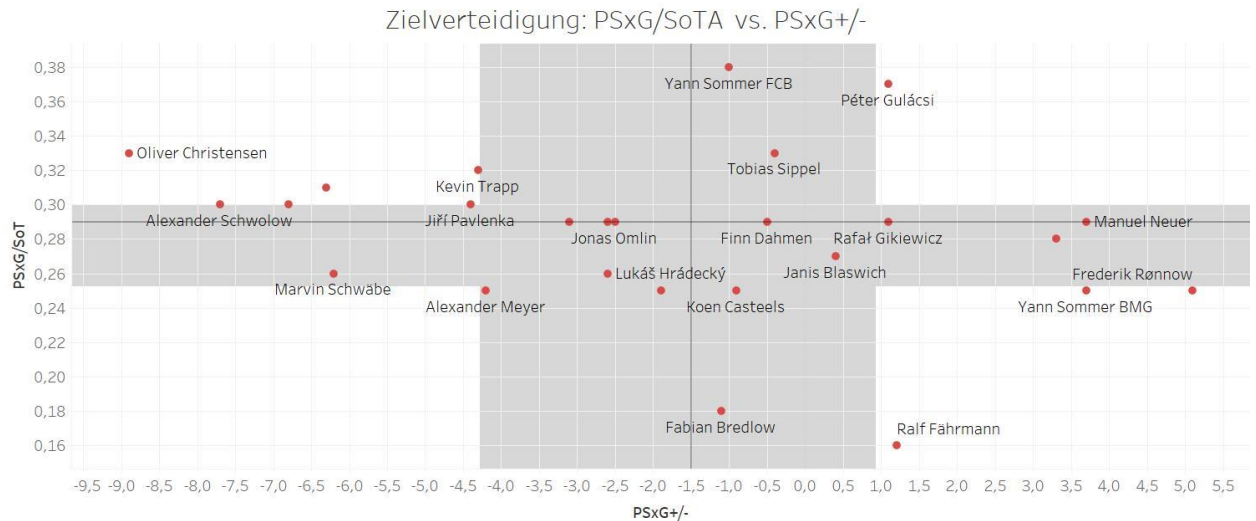


Abbildung 13: PSxG/SoTA vs. PSxG+/- Quelle: eigene Darstellung

Zur Identifikation von Torspielern, die eine geringere Anzahl von Gegentoren verzeichneten, als auf Grundlage eines verhältnismäßig hohen PSxG/90-Wertes erwartet wurde, kann alternativ auch nur der PSxG/90-Wert in Relation zum PSxG+/-/90-Wert gesetzt werden. Es ist jedoch zu beachten, dass dabei die Anzahl der Schüsse auf das Tor unberücksichtigt bleibt, welche ein bedeutender Faktor im jeweiligen Anforderungsprofil darstellen kann (vgl. Abbildung 14).

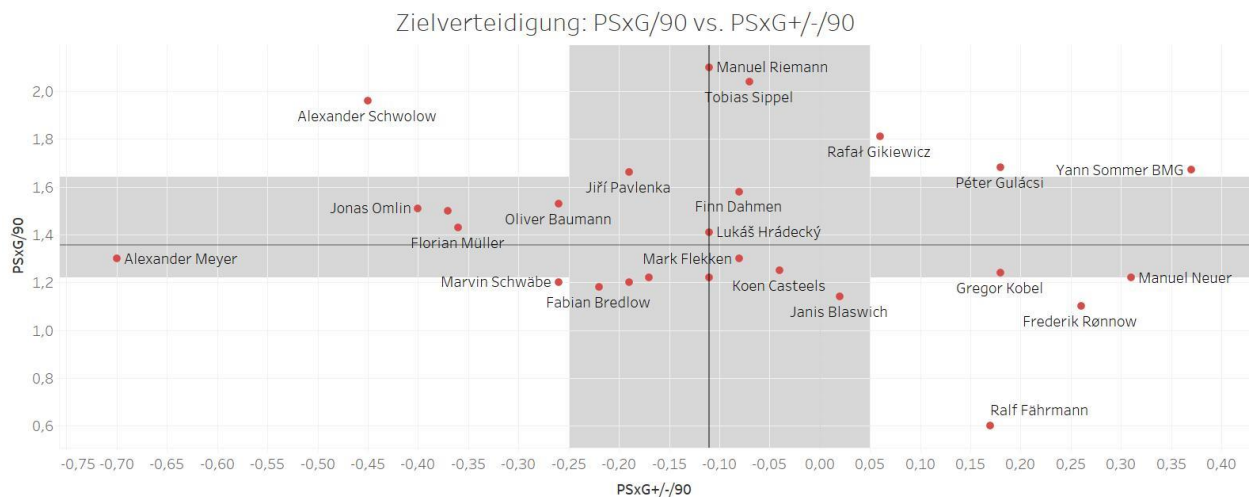


Abbildung 14: PSxG/90 vs. PSxG+/-/90 Quelle: eigene Darstellung

In Abbildung 15 zeigt sich eine deutliche Korrelation zwischen der Fangquote und dem PSxG+/-/90-Wert. Es ist jedoch zu berücksichtigen, dass die Qualität der Torabschlüsse in der Berechnung der Fangquote nicht einbezogen wird, was zu einer Wahrnehmungsverzerrung führen kann. Konkret besteht die Gefahr, dass Torspieler wie Ralf Fährmann und Fabian Bredlow, die zweifellos eine gute Fangquote aufweisen, fälschlicherweise überbewertet werden, während beispielsweise Yann Sommer (FCB) und Peter Gulasci unterbewertet werden. Es ist wichtig zu betonen, dass dies einer der häufigsten Fehler ist, den die Medien (vgl. 5.1) bei der Erstellung Janik Backhaus

von Torhüter-Rankings basierend auf der Fangquote machen.

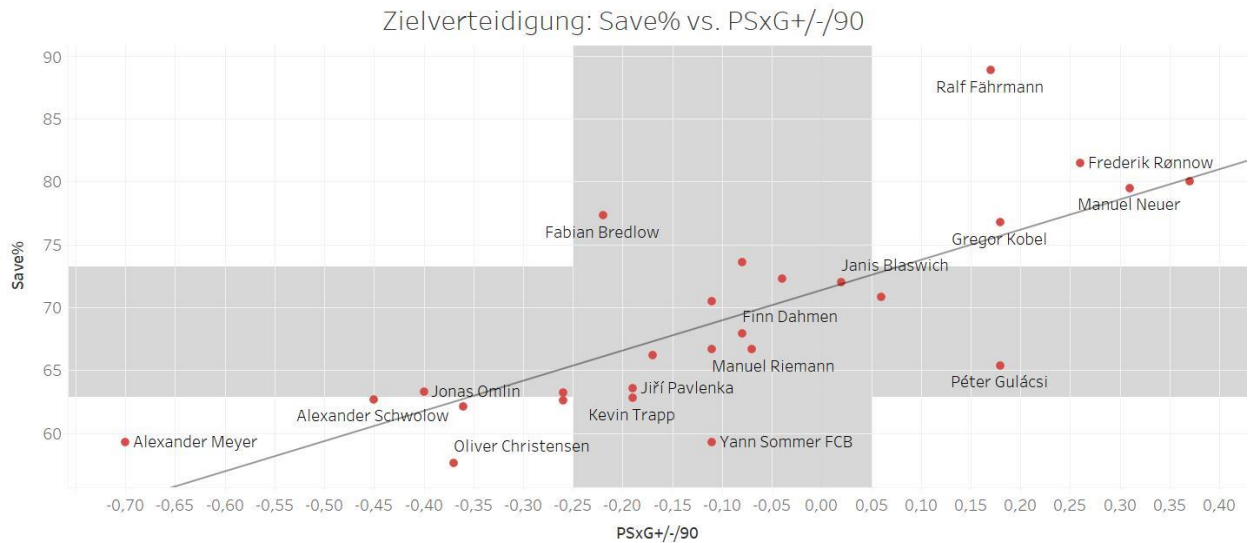


Abbildung 15: Save% vs. PSxG+/-/90 Quelle: eigene Darstellung

Durch die Berücksichtigung der Qualität der gegnerischen Torschüsse in einer adjustierten Fangquote ( $Save\% * (P_{sxG}/SoTA/90)$ ) kann ein realistischeres Bild der tatsächlichen Leistungsfähigkeit im Bereich der Zielverteidigung erlangt werden, während die Korrelation mit etwas größerer Streuung erhalten bleibt (siehe Abb. 16). Ziegler (2021) argumentiert, dass es aus qualitativer Sicht sinnvoll sei, zwischen Aktionen mit Anspruch und solchen ohne Anspruch zu unterscheiden (vgl. 4.1). Es ist zweifellos eine Qualität eines Torspielers, einfache Bälle zu halten, ohne dabei Leichtsinnsfehler zu machen, die zu Gegentoren führen. Jedoch erscheint es im Rahmen der quantitativen Beurteilung seiner Leistungsfähigkeit wesentlich sinnvoller, eine adjustierte Fangquote (wie sie hier vorliegt) als messbaren KPI für die Fähigkeit zu betrachten, schwierige oder unhaltbare Bälle zu parieren und somit den positiven Unterschied für das eigene Team zu machen können.

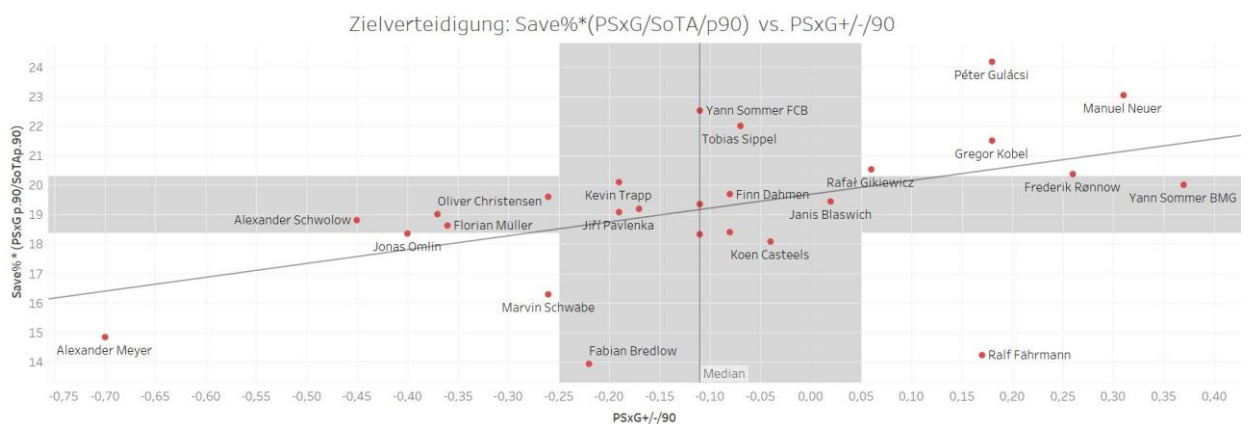


Abbildung 16: adjustierte Fangquote vs. PSxG +/-/90 Quelle: eigene Darstellung

Abbildung 17 zeigt die reine hierarchische Anordnung der Spieler nach ihrer adjustierten Fangquote. Aus Sicht des Autors stellt dies die fairste und am wenigsten verzerrte Darstellungsmöglichkeit der Fangquote dar und bildet den zweiten wesentlichen KPI der Zielverteidigung. Ein reiner Blick auf die Zahlen der adjustierten Fangquote ist nicht so selbsterklärend wie die Rohdaten der Fangquote, die sich immer auf 100% beziehen. Eine Normalisierung der Metrik auf den gleichen Wertebereich würde jedoch aufgrund des traditionellen Begriffsverständnisses der Fangquote eine falsche Vorstellung vermitteln.

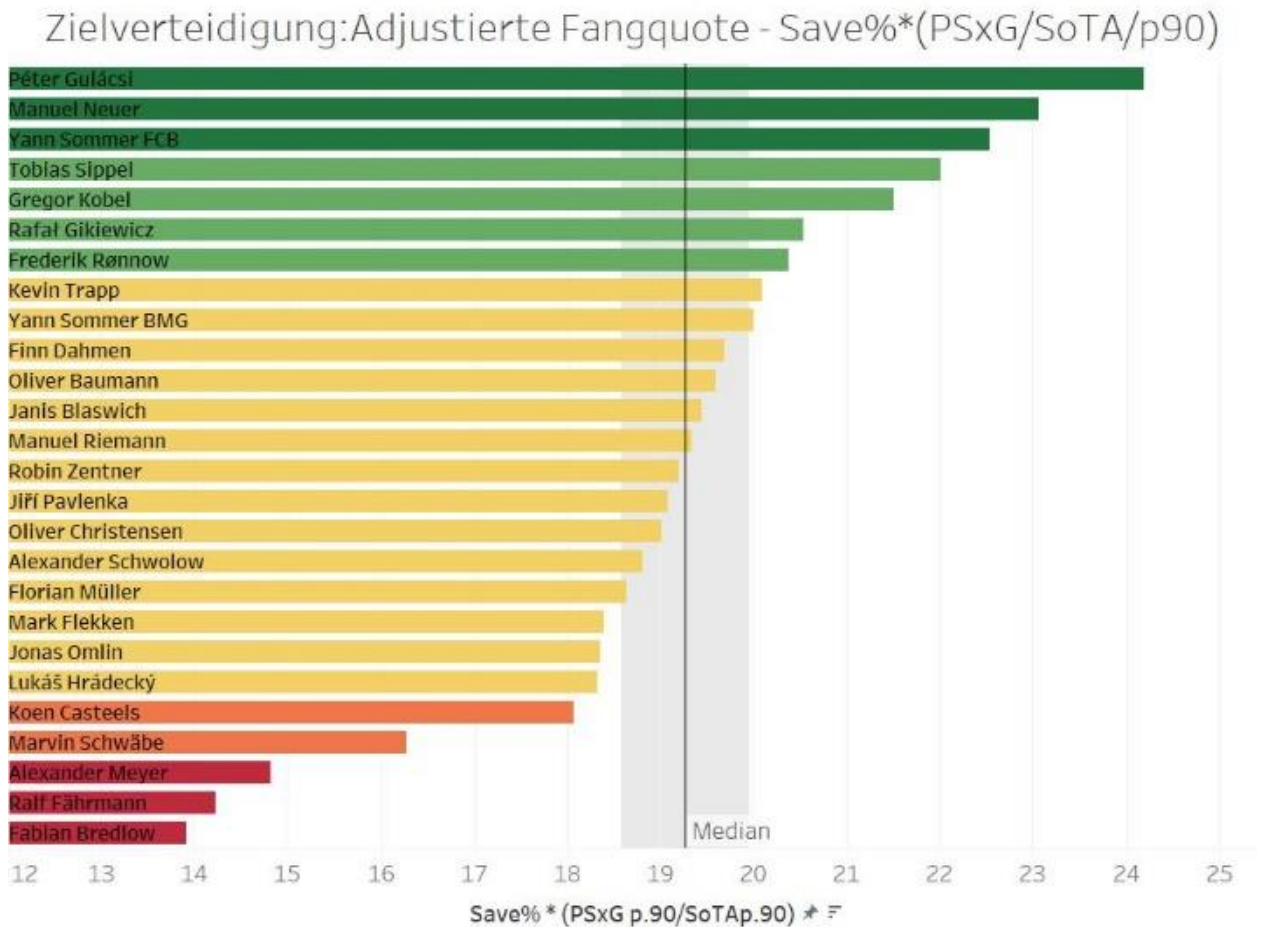


Abbildung 17: adjustierte Fangquote Quelle: eigene Darstellung

Es erscheint hingegen sinnvoll, nicht nur die Schussqualität, sondern auch die absolute Anzahl der Schüsse auf das Tor pro Spiel als eine zweite rein quantitative Variable zu berücksichtigen. Dies hilft insbesondere dabei, jene Torspieler zu identifizieren, die trotz der Konfrontation mit sehr wenigen und tendenziell gefährlicheren gegnerischen Torschüssen eine sehr gute Fangquote aufweisen. Denn dies ist im Bereich der Zielverteidigung die Kernkompetenz von Torspielern in Spitzenmannschaften. Eine solche, zweite Anpassung der Fangquote wurde in Abbildung 18 dargestellt. Die farbliche Abstufung der Leistungsfähigkeit aus Abbildung 17 (angepasste Fangquote auf Basis der Qualität der Torschüsse) wurde beibehalten, um die Verschiebungen der einzelnen Spieler im direkten Vergleich zu verdeutlichen.

Zielverteidigung: Ad. Fq. II -  $(\text{Save}\% * (\text{PSxG}/\text{SoTA}/\text{p90})) / \text{SoTA}/90$

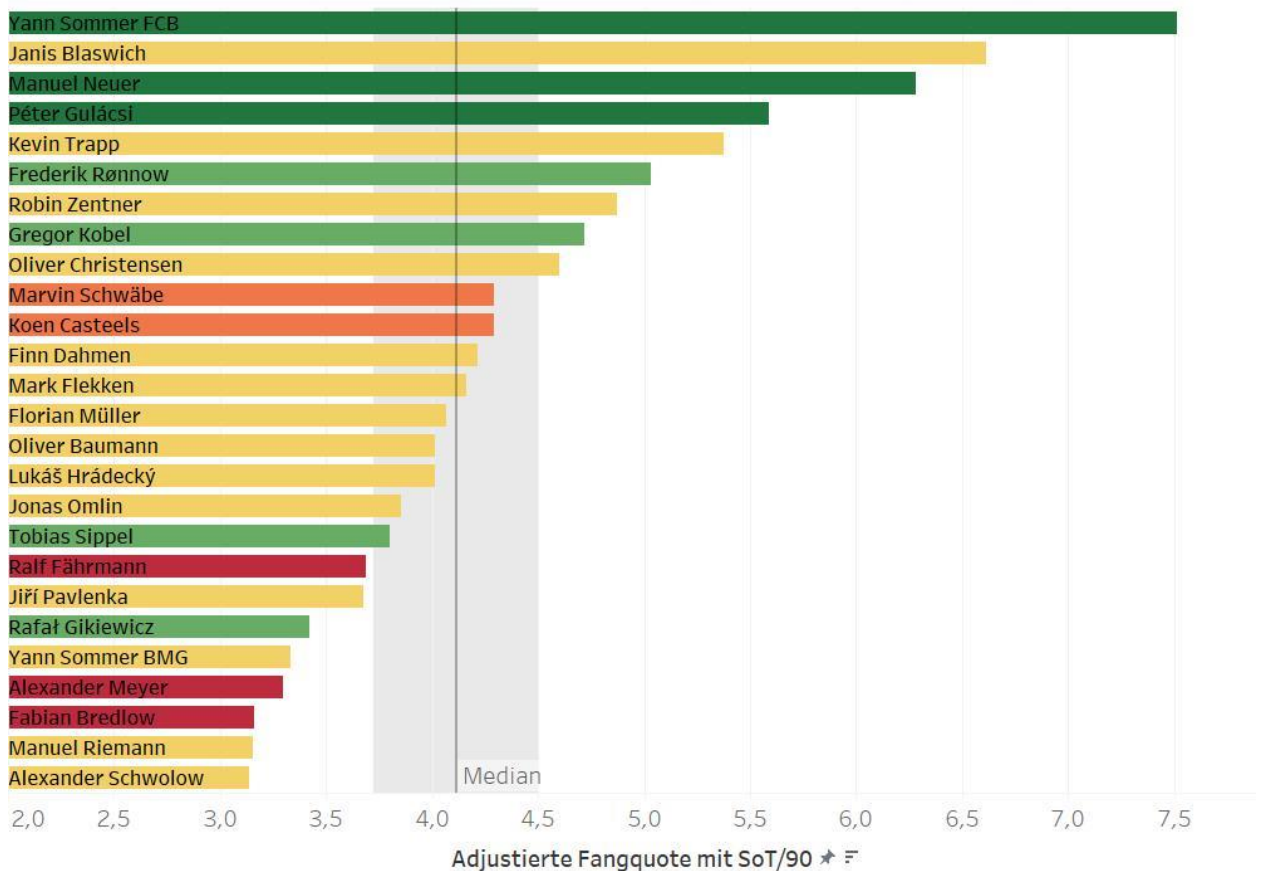


Abbildung 18: Adjustierte Fangquote II mit Berücksichtigung von SoTA/90 Quelle: eigene Darstellung

Der interessierte Leser mit rudimentärer qualitativer Beurteilungsfähigkeit und Kenntnissen der Bundesliga sollte spätestens an diesem Punkt erkennen, warum es notwendig ist, die Rohdaten auf quantitativer Ebene in den Kontext des Spiels zu setzen und verstehen, dass die vorherigen Analyseschritte unzureichend waren. Ein deutliches Beispiel hierfür ist Nationalspieler Kevin Trapp, der, obwohl er in dieser Bundesligasaison sicherlich nicht in Bestform spielt, in keiner der vorherigen Analyseschritte in dem Bereich angesiedelt war, wo man ihn aufgrund der qualitativen Eindrücke seiner Leistung erwartet hätte. Auf dieser Datenebene liegt er jedoch wieder im Bereich der Top-5. Ein weiteres Beispiel für die Notwendigkeit der spielkontextbezogenen Adjustierung von datenbasierten Leistungsbeurteilungen ist der Vergleich der Leistungen von Yann Sommer. Während er bei Borussia Mönchengladbach vielen, tendenziell einfacheren Abschlüssen gegenüberstand und dementsprechend oft glänzte, ist er bei Bayern München mit wenigen, aber schwierigeren Abschlüssen konfrontiert. Um seine Leistungsfähigkeit im Bereich der Zielverteidigung auf Datenbasis aussagekräftig beurteilen zu können, ist eine spielkontextbezogene Adjustierung unerlässlich.

Abschließens soll betrachtet werden, wie die "Clean Sheet-Quote" bzw. die Anzahl der "Weißen Westen" objektiver genutzt werden kann. Es ist nicht sinnvoll, diese isoliert und ohne Kontext als Qualitätskriterium für Torspieler zu betrachten. Ob es gelingt "zu Null" zu spielen, hängt stark davon ab, wie gut die eigene Mannschaft verteidigt bzw. wie viele Torschüsse des Gegners zugelassen werden. Diese Zusammenhänge werden durch die Korrelation zwischen den SoTA/90 und der Anzahl der CS/90 verdeutlicht, welche trotz der geringen Stichprobengröße bereits in dieser Arbeit in Ansätzen erkennbar ist. Im Rahmen des Scoutings ist daher Abbildung

19 von großer Relevanz:

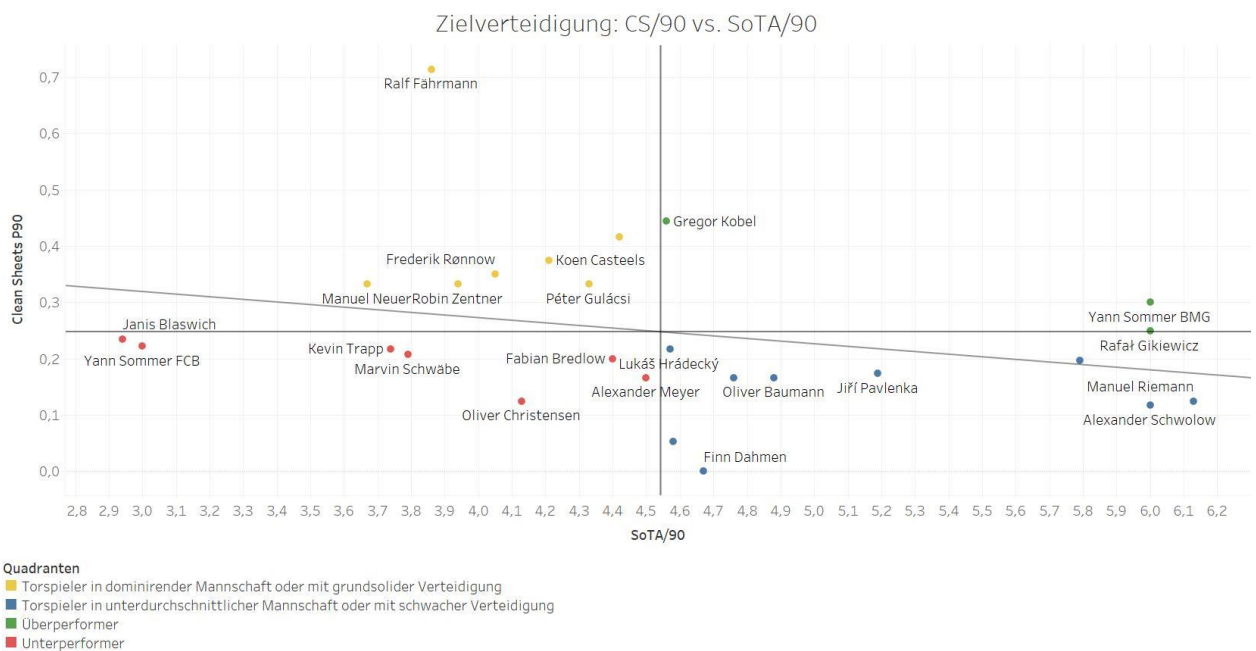


Abbildung 19: Clean Sheets/90 vs. SoTA/90 Quelle: eigene Darstellung

Die linke obere Ecke wird von Torspielern (gelb) besetzt, die mit einer soliden Abwehr und/oder in dominanten Mannschaften spielen und daher mit wenigen Torschüssen konfrontiert werden und eine hohe Anzahl von zu Null-Spielen aufweisen. In der rechten unteren Ecke befinden sich Torspieler (blau), die in unterdurchschnittlichen Mannschaften oder mit einer schlechten Abwehr spielen und daher viele gegnerische Torschüsse zulassen, was zu einer niedrigen Clean Sheet-Quote führt. Die beiden Quadranten, die hier von Interesse sind, sind die in der oberen rechten Ecke und in der unteren linken Ecke des Diagramms. Die Überperformer werden durch grüne Markierungen gekennzeichnet und sind jene Torspieler, die trotz überdurchschnittlich vieler Torschüsse (>4,54 P90) überdurchschnittlich viele zu Null-Spiele (>0,24 P90) erzielen können. Die Unterperformer sind am anderen Ende des Spektrums und werden durch rote Markierungen gekennzeichnet. Neben der Berücksichtigung der Anzahl der Schüsse könnte zusätzlich eine qualitative Bewertung der Torabschlüsse herangezogen werden. Jedoch führt dies lediglich zu



geringfügigen Verschiebungen zwischen den Clustern, wie in Abbildung 20 dargestellt.

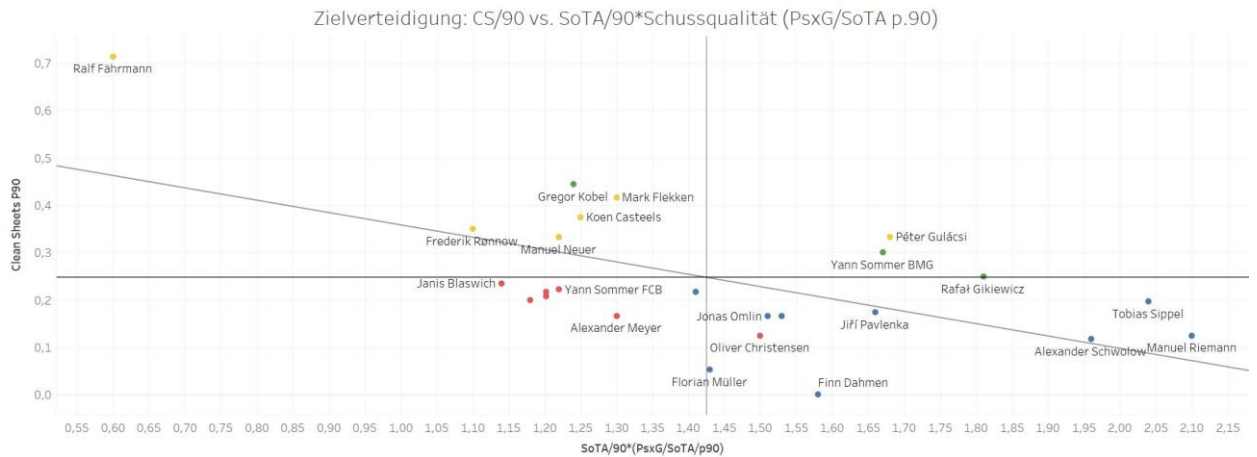


Abbildung 20: Clean Sheets/90 vs SoTA/90\* Schussqualität Quelle: eigene Darstellung

### 6.1.4 Detailanalyse Mitspielverhalten bei gegnerischem Ballbesitz

Die Analyse des Mitspielverhaltens bei gegnerischem Ballbesitzes wird durch die Betrachtung des zugehörigen KPI's eingeleitet, wie in Abbildung 21 veranschaulicht wird.

#### MSPV BBG: KPI Mitspielverhalten bei gegnerischem Ballbesitz

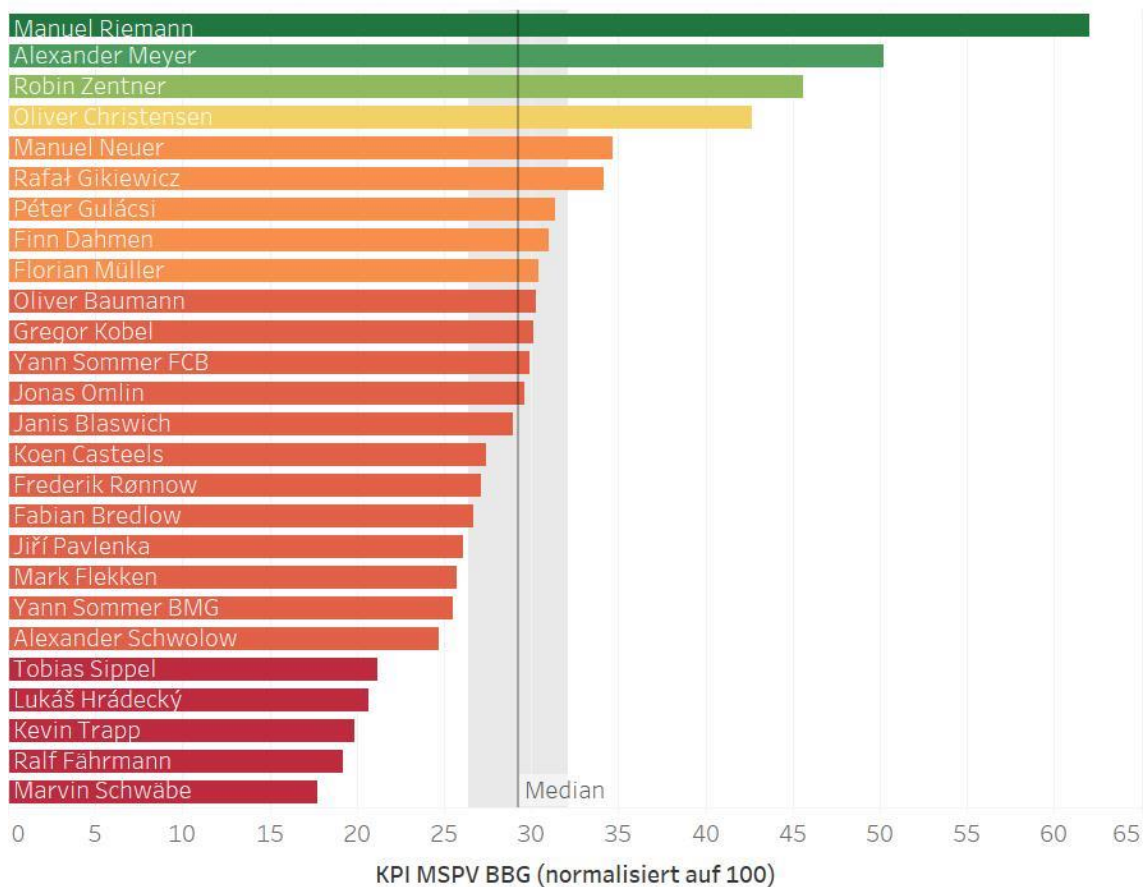
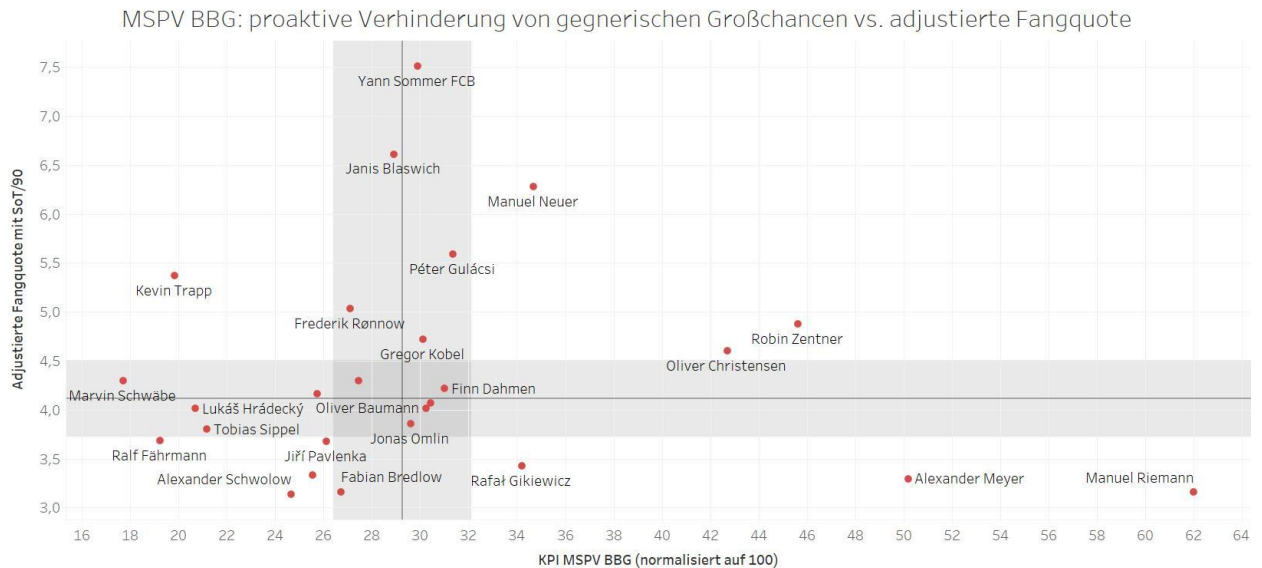


Abbildung 21: KPI Mitspielverhalten bei gegnerischem Ballbesitz Quelle: eigene Darstellung

Im nächsten Schritt wird dieser KPI-Wert in Relation zur adjustierten Fangquote (vgl. Abb.18) gesetzt. Ziel ist es, jene Torspieler zu identifizieren, die durch proaktives Mitspielverhalten in der Lage sind, gegnerische Großchancen zu verhindern, aber auch eine gute Paradenquote aufweisen, wenn diese dennoch entstehen. Diese kategorienübergreifende Analyse wird in Abbildung 22 veranschaulicht.



**Abbildung 22:** proaktive Verhinderung von gegnerischen Großchancen vs. adjustierte Fangquote Quelle: eigene Darstellung

Im nächsten Schritt wird die Proaktivität in Bezug auf die Verhinderung gegnerischer Torchancen durch aktives Mitspielverhalten bei gegnerischen Ballbesitz genauer analysiert. Hierbei wird die durchschnittliche Distanz der defensiven Aktionen in Relation zur Anzahl der defensiven Aktionen außerhalb des eigenen Strafraums gesetzt. Es ist anzumerken, dass zwischen diesen beiden Werten eine hohe Korrelation besteht, da sie sich gegenseitig bedingen, wie in Abbildung 23 dargestellt wird.

MSPV BBG: Durchschnittliche Distanz von defensiven Aktionen vs. def. Aktionen außerhalb des 16ers/90

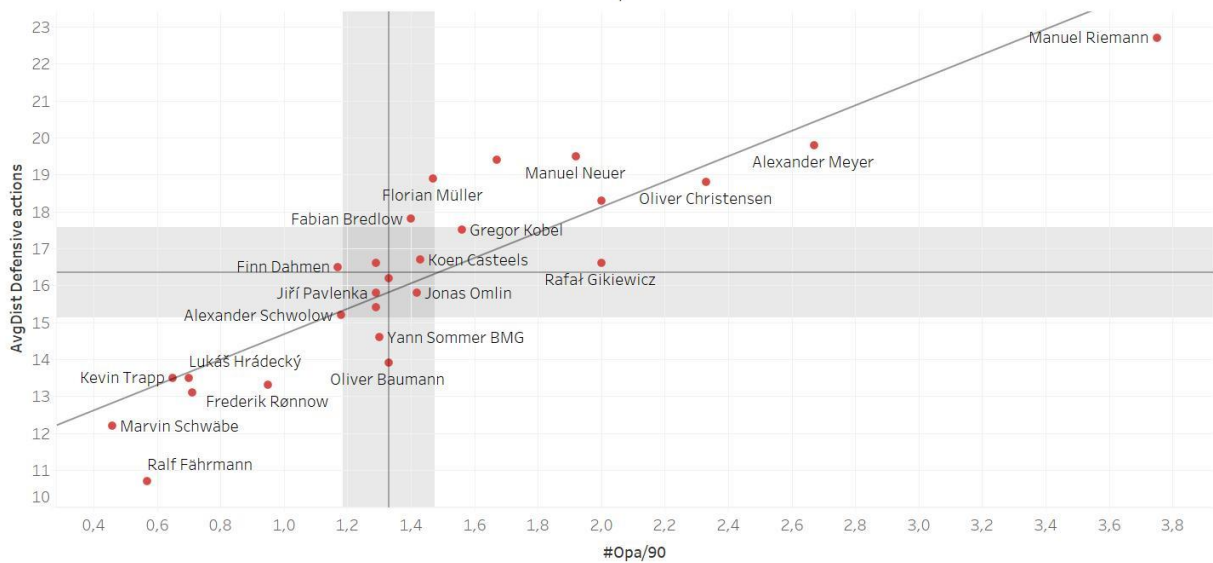


Abbildung 23: durchschnittliche Distanz der defensiven Aktionen vs. #OPA/90 Quelle: eigene Darstellung

Die Platzierung der Spieler in den Quadranten sowie oberhalb und unterhalb der Trendlinie aus Sicht des Scouting ist von Interesse. Aus analytischer Sicht ist der Median der durchschnittlichen Distanz der defensiven Aktionen von Bedeutung. Dieser liegt bei 16,35 Metern und befindet sich somit relativ genau an der Strafraumgrenze. Es ist somit festzustellen, dass die Hälfte aller Bundesliga-Torspieler mit mehr als 20% Einsatzzeit ihren eigenen Strafraum im Ballgewinnspiel kaum verlassen. Dies weist auf Defizite in der Ausbildung, unzureichendes Training oder eine Tendenz zu geringem Mut bei der Entscheidungsfindung und Risikoabwägung hin.

Nach Thaler (2019) ist die Unterstützung der eigenen Mitspieler im Ballgewinnspiel ein zentrales Unterscheidungskriterium zwischen Torspieler 2.0 und Torspieler 3.0 (vgl. 4.1). Ein Blick auf die Zahlen verdeutlicht, dass Spieler, die links unten platziert sind, zum Teil deutlich weniger als eine Aktion außerhalb des eigenen Strafraums pro Spiel durchführen und eine durchschnittliche Distanz ihrer Aktionen von weniger als 13,5 Metern haben. Sie können daher auch noch im Jahr 2023 eher als Torspieler 2.0 oder Linientorhüter charakterisiert werden.

Es ist zu beachten, dass beide Werte wesentlich von der Höhe der eigenen Kette abhängen und insbesondere durch das mannschaftstaktische Verhalten im Umschalten nach Ballverlust beeinflusst werden. Einige Cheftrainer verfolgen offensichtlich taktische Ansätze, die auch in der heutigen Zeit noch eine gewisse Daseinsberechtigung für diesen Torspieler Typ 2.0 ermöglichen. Es sei darauf hingewiesen, dass diese Feststellung keine Wertung der der jeweiligen

Spielphilosophien beinhaltet und die Abbildung keine qualitative Bewertung der Torspieler liefert, sondern lediglich dazu dient, das Spielverhalten genauer zu charakterisieren.

Grundsätzlich können die Geschwindigkeit sowie die Anzahl der Sprints und intensiven Läufe gute Indikatoren dafür sein, um die Fähigkeit eines Torspielers zum Ablaufen von Bällen hinter die Kette zu beurteilen (vgl. Abbildung 24). Es ist jedoch zu beachten, dass die Geschwindigkeit verzerrt sein kann, da der Torspieler möglicherweise in seinen Einsätzen nicht mit einer Situation konfrontiert wurde, in der er seine maximale Geschwindigkeit erreichen musste. Die Anzahl der Sprints und intensiven Läufe hängt natürlich auch davon ab, wie oft der Spielkontext es erfordert hat, dass der Torspieler in diesem Geschwindigkeitsbereich läuft. Obwohl die vorliegende Analyse ausschließlich auf „kontextlosen Trackingdaten“ beruht und somit sowohl die Entscheidungsfindung als auch die Position des Torspielers auf dem Spielfeld in Bezug auf den Ball und seine Mitspieler vollständig unberücksichtigt lässt, kann durch den Vergleich der Spielerpositionierungen in den Abbildungen 23 & 24 eine hohe Validität angenommen werden.

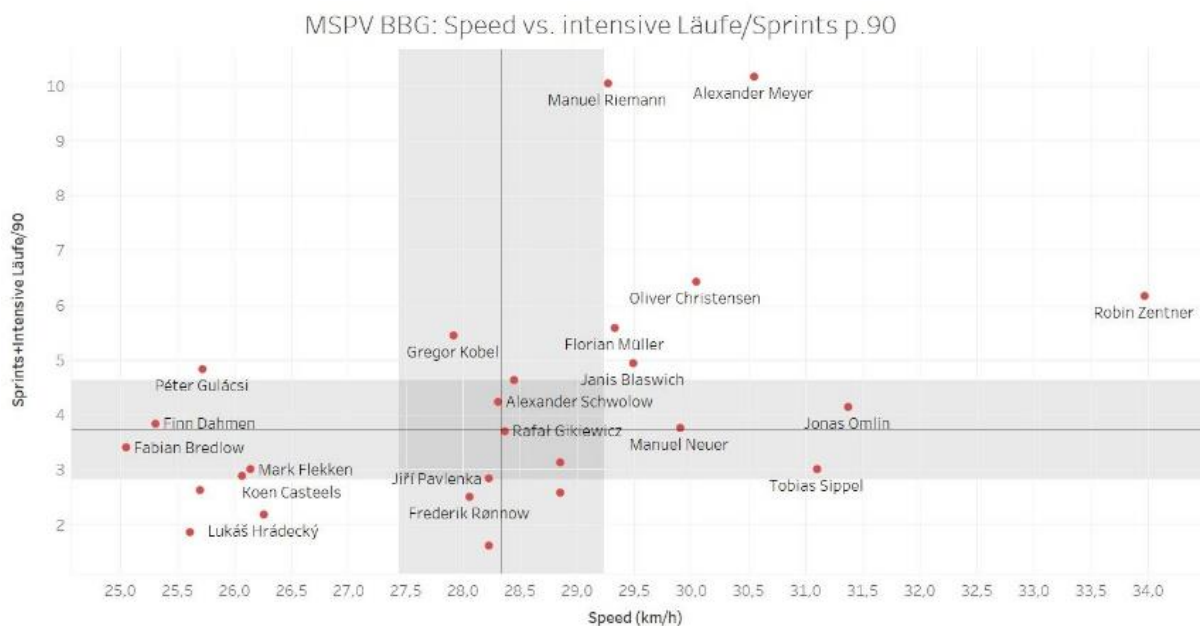


Abbildung 24: Speed vs. Sprints + intensive Läufe/90 Quelle: eigene Darstellung

Als Ergänzung zu den bisherigen Erkenntnissen erscheint es sinnvoll, die Anzahl von Fehlern in den mitspielenden Aktionen, die zu gegnerischen Torschüssen führen, zu berücksichtigen (siehe Abbildung 25). Um die Übersichtlichkeit in dieser ergänzenden Abbildung zu wahren, wurden Spieler ausgeschlossen, die bisher keine derartigen Fehler begangen haben oder bei denen

keine entsprechenden Aktionen erfasst wurden. Diese Spieler sind für das Scouting besonders interessant.

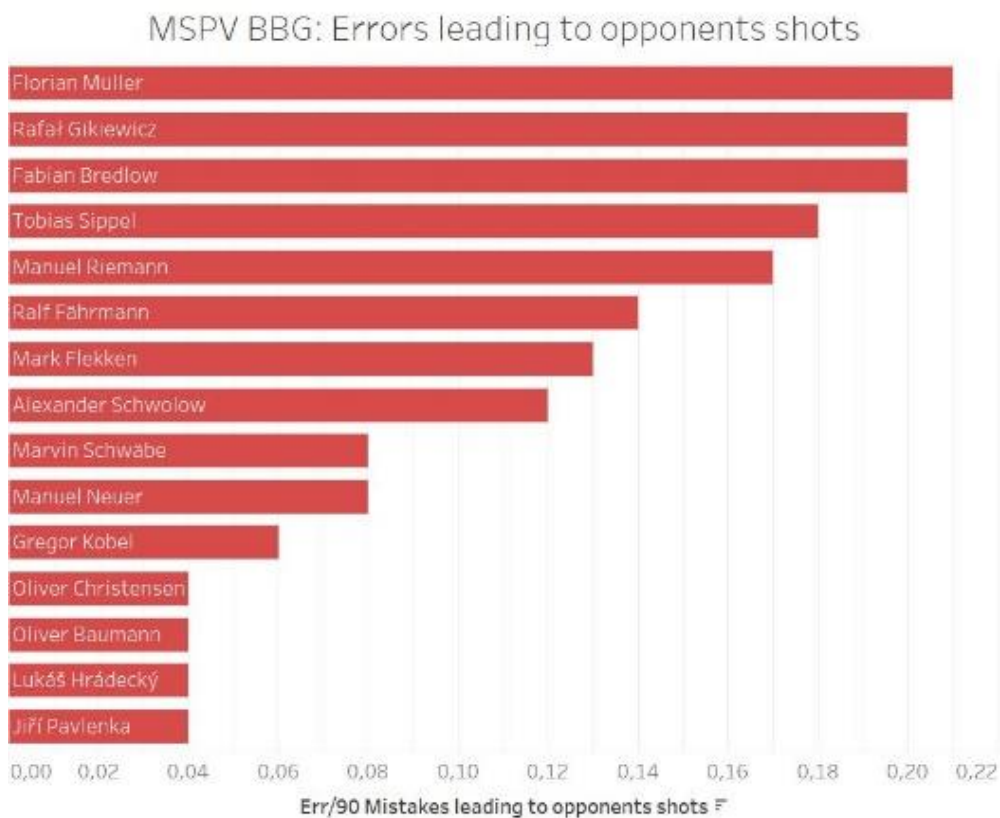


Abbildung 25: Errors leading to opponents shots Quelle: eigene Darstellung

Aus Perspektive des Torspielerscoutings ist eine Untersuchung der feldspielerspezifischen Ereignisse Clearances, Challenges, Tacklings und Interceptions nicht aussagekräftig, da viele artgleiche Torspieler-Aktionen mit verwendeten Definitionskatalog offensichtlich nicht erfasst werden. Stattdessen werden diese Aktionen lediglich als #OPA-Aktionen, dem aktionsübergreifenden torspielerspezifischen Äquivalent, in die Statistik aufgenommen, ohne dass eine genaue Differenzierung vorgenommen wird. Dies führt zu unerwartet niedrigen Werten bei den genannten Ereignissen und macht eine Analyse dieser Werte sinnlos.

Zur Quantifizierung der Strafraumbeherrschung als zweiten wesentlichen Teilbereich des Mitspielerverhaltens bei gegnerischem Ballbesitz bietet der vorhandene Datensatz nur begrenzte Möglichkeiten. Eine genauere Differenzierung zwischen verschiedenen spielkontextrelevanten Faktoren wie der Art der Flanken (z.B. hoch, halbhoch oder flach), dem Zielraum in Bezug auf die Torspielerposition, der Spielsituation (z.B. aus dem Spiel oder nach Standards) sowie der spezifischen Torspieler-Aktion (z.B. Fangen, Ablenken oder Fausten) und die Unterscheidung, ob der Torspieler versucht hat, den Ball abzufangen oder sich entschieden hat auf der Linie zu bleiben, wäre wünschenswert. Aufgrund der Einschränkungen des verwendeten Datensatzes ist

eine solche differenzierte Analyse jedoch nicht möglich. Aus Abbildung 26 lassen sich somit lediglich die Anzahl der zugelassenen Flanken seiner Mannschaft pro 90 Minuten sowie die prozentuale Quote der abgefangenen Flanken ableiten.

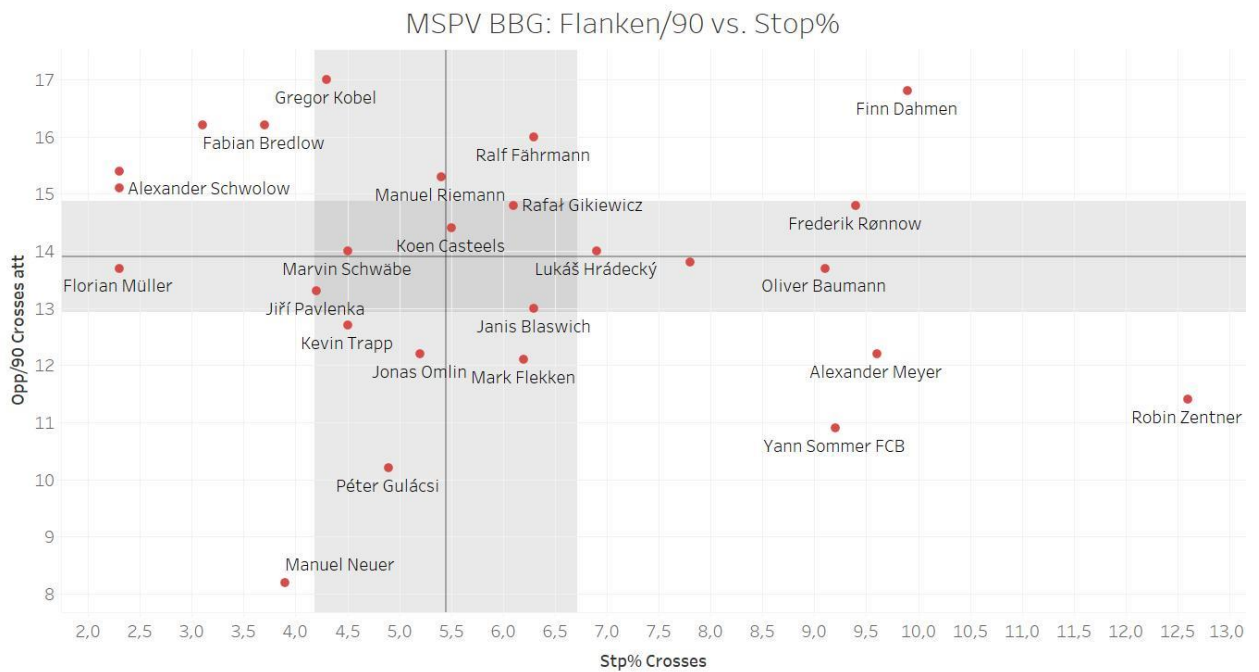


Abbildung 26: Flanken/90 vs. Stop % Quelle: eigene Darstellung

### 6.1.5 Detailanalyse Mitspielverhalten im eigenen Ballbesitz

Für die Analyse des Mitspielverhaltens im eigenen Ballbesitz bietet sich als Einstieg erneut die hierarchische Reihenfolge des selbstberechneten Key Performance Indicators an, welche in Abbildung 27 dargestellt ist.

### MSPV BB: KPI Mitspielverhalten im eigenen Ballbesitz

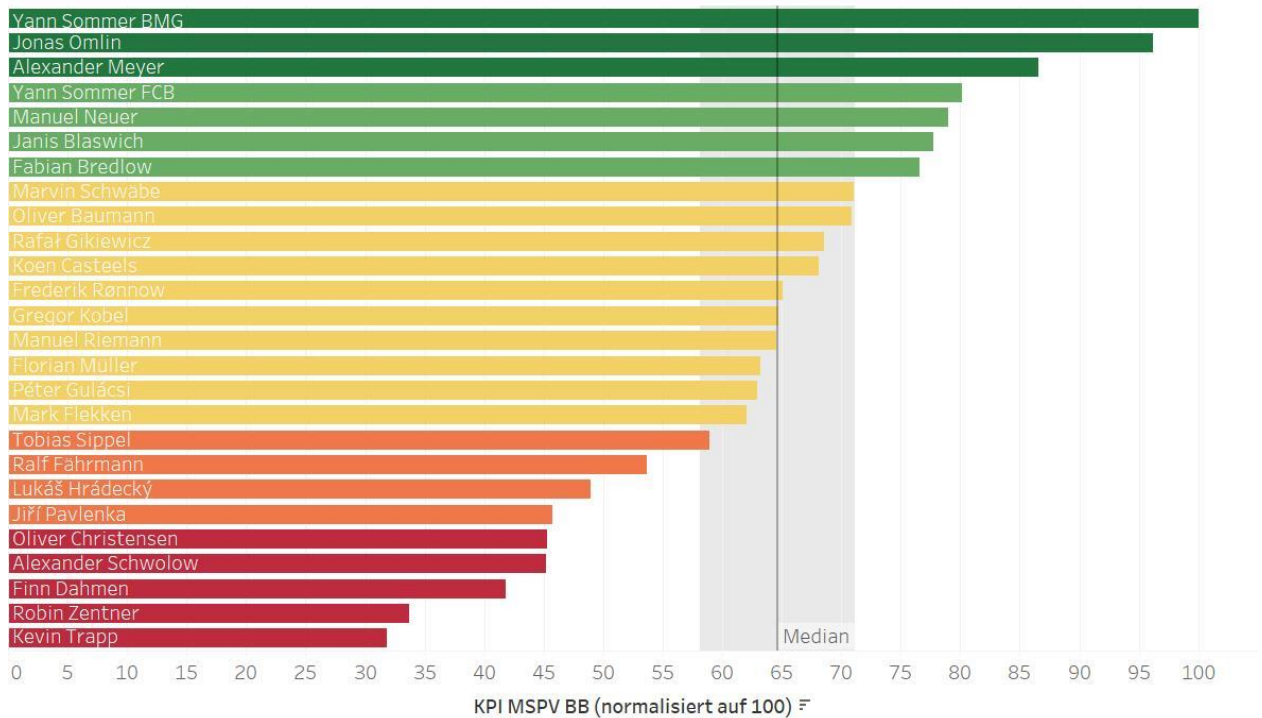


Abbildung 27: KPI Mitspielverhalten im eigenen Ballbesitz Quelle: eigene Darstellung

Im Rahmen der detaillierten Analyse erfolgt eine grundlegende Unterscheidung zwischen Pässen und Ballkontakten, welche in der Praxis eng miteinander verbunden sind. Es ist jedoch zu berücksichtigen, dass es im Rahmen dieser Studie nicht möglich war, die absoluten Zahlen mit Ballbesitz-Metriken auf Mannschaftsebene zu bereinigen, um einen noch gerechteren Vergleich der individuellen Spielerleistungen zu ermöglichen (vgl. 5.2). Zunächst wird den Pässen besondere Aufmerksamkeit geschenkt: Mittels einer gegenüberstellenden Analyse der Anzahl erfolgreicher Pässe und der Passquote, wie in Abbildung 28 dargestellt, werden zunächst diejenigen Torspieler identifiziert, die viele erfolgreiche Pässe verzeichnen und eine Tendenz zu einer geringen Anzahl von Fehlpässen aufweisen.

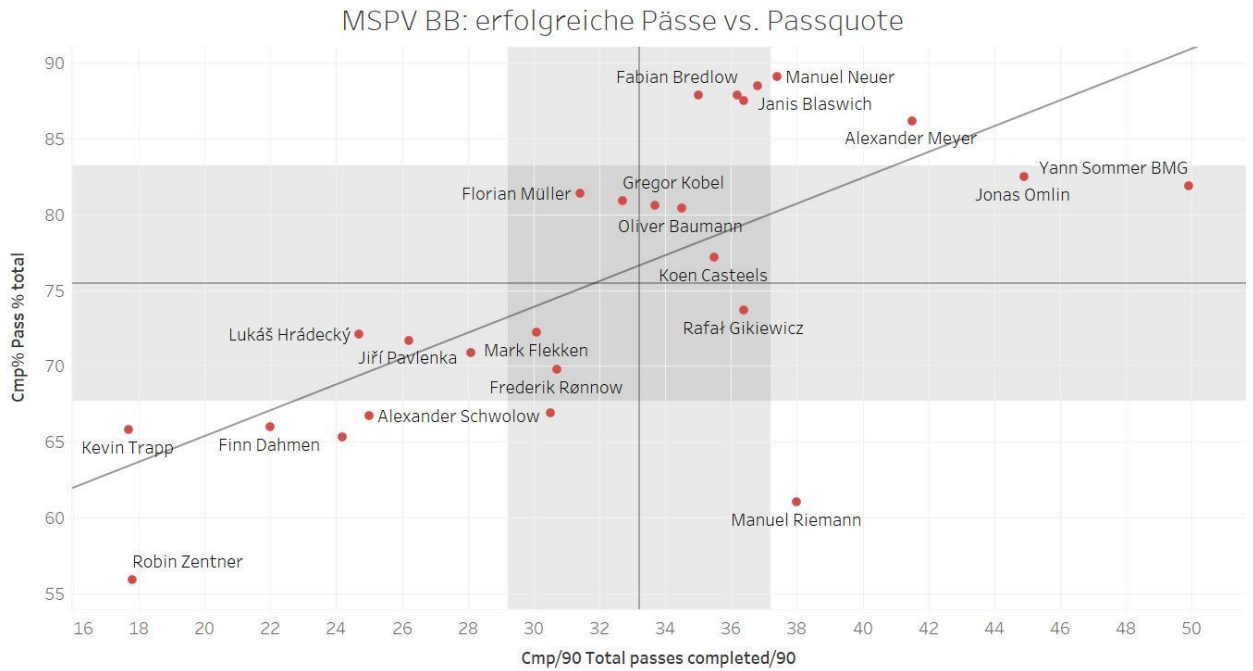


Abbildung 28: erfolgreiche Pässe/90 vs. Passquote Quelle: eigene Darstellung

Die vorliegende Analyse der Anzahl erfolgreicher Pässe und der Passquote weist eine begrenzte Aussagekraft auf, obwohl eine Korrelation zwischen beiden Werten offensichtlich ist. Ein entscheidender Faktor, der die Passquote beeinflusst, ist die Passlänge, da längere Risikopässe naturgemäß ein höheres Ballverlustrisiko mit sich bringen. Einige Trainer nutzen dieses Risiko als bewusstes taktisches Mittel, beispielsweise durch Spiel auf den zweiten Ball bzw. sofortiges Gegenpressing nach einem Ballverlust nahe am gegnerischen 16er. In Abbildung 29 wird die durchschnittliche Passlänge im Verhältnis zur Anzahl der Passversuche dargestellt, wobei die Passquote farblich gekennzeichnet ist. Grundsätzlich ist wie erwartet zu erkennen, dass kürzere Pässe zu einer höheren Erfolgsquote führen. Die Interpretation der Grafik und die Identifikation interessanter Spieler hängen wesentlich davon ab, ob kurze oder lange Bälle im Spielaufbau bevorzugt werden. Aus Scouting-Perspektive wären insbesondere diejenigen Spieler von Interesse, die trotz vieler, tendenziell längerer Bälle eine vergleichsweise hohe Passquote verzeichnen. Allerdings gibt es nur einen einzigen solchen Spieler, Koen Castels, und die Differenz ist nicht signifikant. Es konnten zudem keine negativen Ausreißer in die andere Richtung identifiziert werden, das heißt Spieler mit einer schlechten Passquote trotz geringer Anzahl kurzer Pässe.



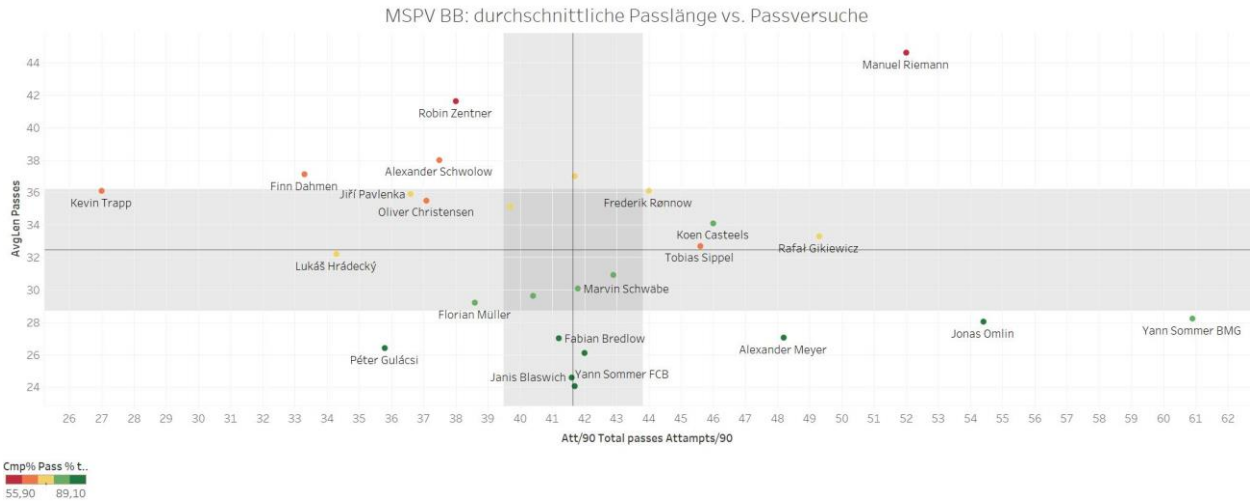


Abbildung 29: durchschnittliche Passlänge vs. Passversuche/90 vs. Passquote Quelle: eigene Darstellung

Abbildung 30 bietet einen detaillierteren Einblick in die Passquote in Bezug auf unterschiedliche Passlängen. Die totale Passquote ist in Blau dargestellt, während kurze Pässe (5-15 Meter) in Grün, mittlere Pässe (15-30 Meter) in Gelb und lange Pässe (mehr als 30 Meter) in Rot dargestellt sind. Interessanterweise scheint eine Passquote nahe der 100% bei kurzen und mittleren Pässen eine grundlegende Anforderung für das Spielen auf diesem Niveau zu sein, weshalb diese Werte in Bezug auf Scouting von eher untergeordneter Bedeutung sind.

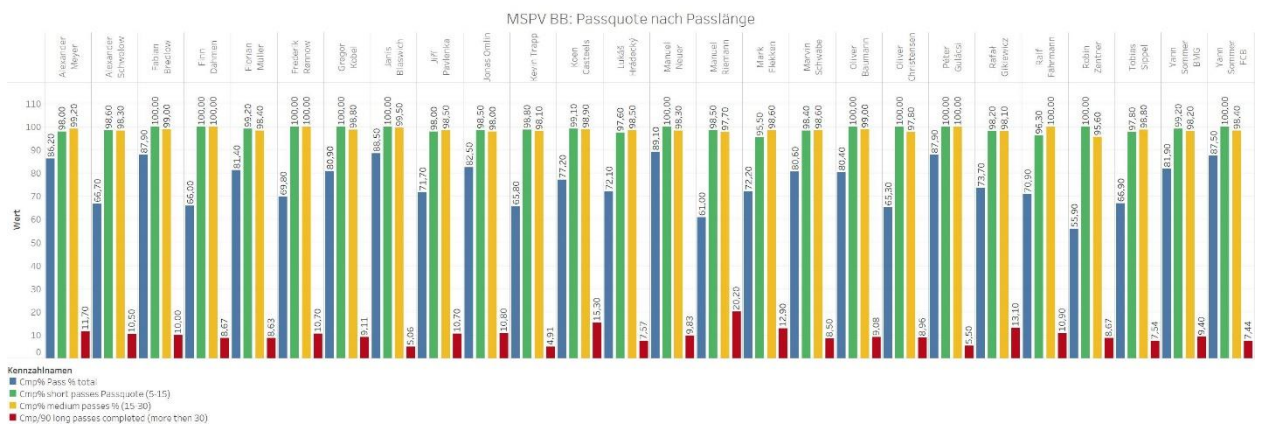


Abbildung 30: Passquote nach Passlänge Quelle: eigene Darstellung

Die Abbildungen 31 und 32 geben Aufschluss über die präferierte Passlänge der Torspieler im Spielaufbau. Hierbei wird die Gesamtzahl erfolgreicher Pässe bzw. die Anzahl der Passversuche in die genannten Passlängen aufgeteilt. Dadurch wird ein relativ genauer Überblick über die individuelle Passlängenverteilung jedes Torspielers ermöglicht.

### MSPV BB: erfolgreiche Pässe nach Passlänge

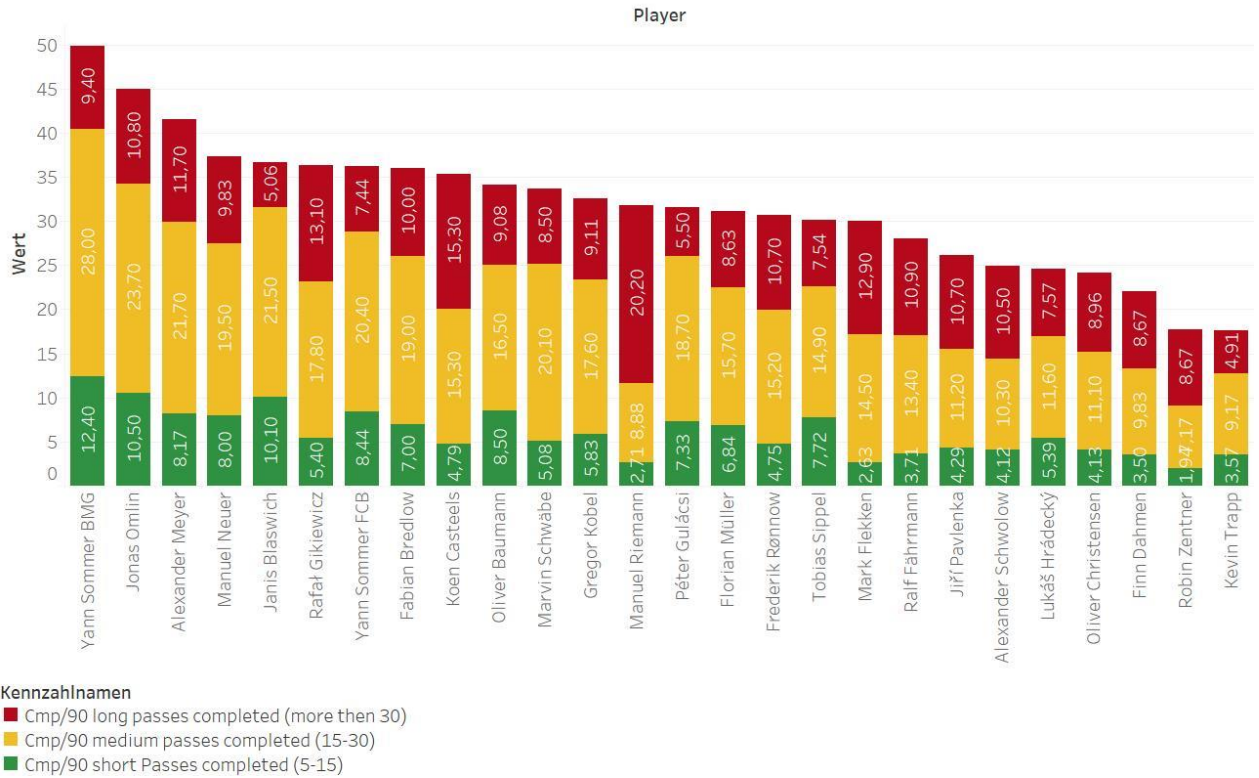


Abbildung 31: erfolgreiche Pässe nach Passlänge Quelle: eigene Darstellung

### MSPV BB: Passversuche nach Passlänge

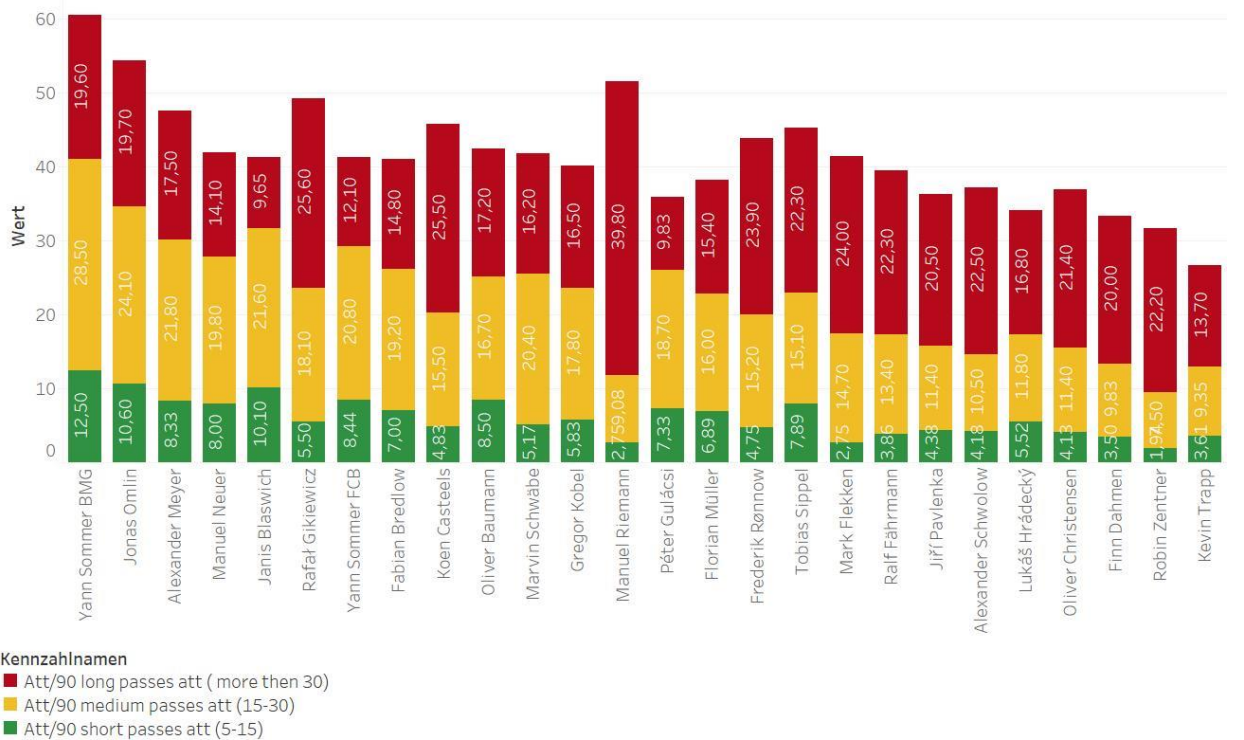


Abbildung 32: Passversuche nach Passlänge Quelle: eigene Darstellung

Bei der Bestimmung der präferierten Passlänge ist zu berücksichtigen, dass diese maßgeblich von der Spielidee der jeweiligen Mannschaft im Spielaufbau abhängt. Im Rahmen des Scouting-Prozesses ist es wichtig zu beurteilen, inwiefern ein Spieler in das eigene Team integriert werden kann. Spieler, die eine Vorliebe für lange Bälle zeigen, könnten Schwierigkeiten haben, sich in ballbesitzorientierten Mannschaften zurechtzufinden, die einen kontrollierten Spielaufbau mit kurzen und mittleren Pässen bevorzugen. Grund dafür ist, dass diese nicht an das sichere Kurzpassspiel unter Raum-, Zeit- und Gegnerdruck gewöhnt sind. Der umgekehrte Prozess, bei dem Spieler, die das sichere Kurzpassspiel beherrschen, aufgefordert werden gezielte lange Bälle spielen, wenn die jeweilige Spielsituation oder Spielphilosophie dies erfordert, erscheint aus praktischer Erfahrung heraus einfacher. Das ideale Passprofil eines Torspielers zeichnet sich durch die Fähigkeit aus, sowohl den Ballbesitz mit vielen kurzen und mittleren Pässen zu sichern, um den Gegner zu locken und somit Lücken zwischen, innerhalb und hinter den gegnerischen Linien zu schaffen, als auch gezielte lange Vertikal- oder Diagonalpässe zu spielen, um diese Lücken auszunutzen.

Im nächsten Schritt wird die aktive Einbindung der Torspieler in das Angriffsspiel ihrer Mannschaft beurteilt. Hierfür bietet sich die Verwendung eines Scatterplots an, in dem Livepässe (im Spiel) und sogenannte "Dead Passes", also Pässe nach Standardsituationen, separat dargestellt werden (vgl. Abbildung 33). Dabei ist die Passquote der einzelnen Spieler erneut farblich markiert.

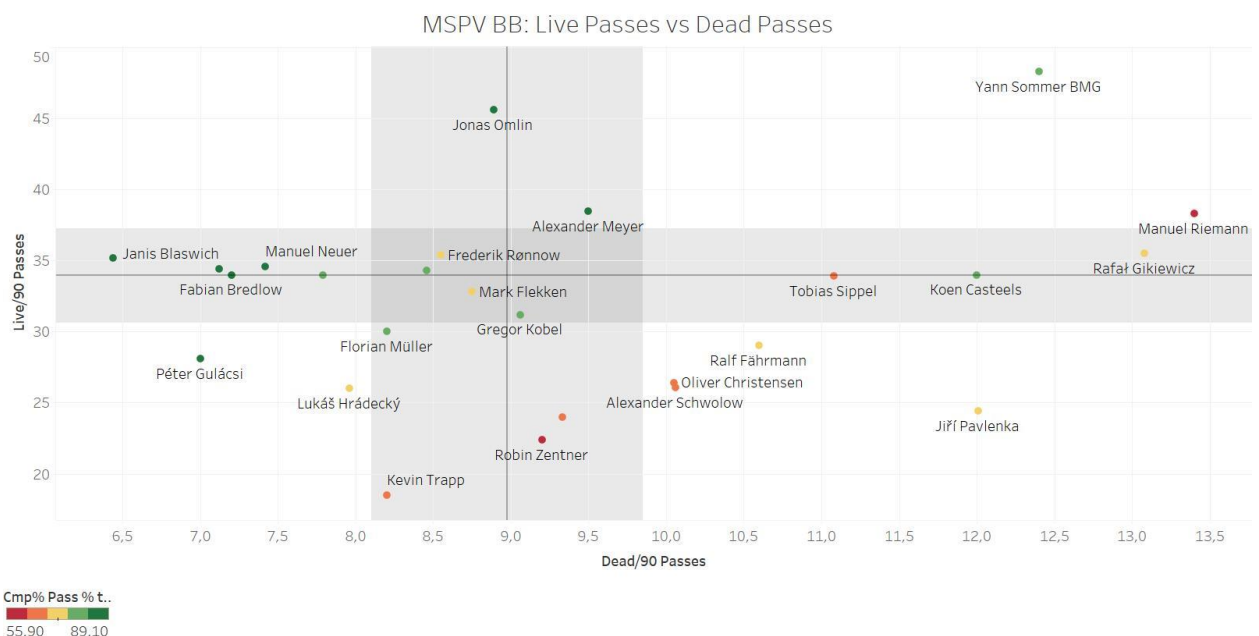


Abbildung 33: Einbindung der Torspieler ins Angriffsspiel Quelle: eigene Darstellung

Obwohl der Bezug zum Ballbesitzwert bzw. das Verhältnis zur Gesamtzahl der Pässe der jeweiligen Mannschaft fehlt, lassen die Spieler im unteren Teil der Grafik vermuten, dass sie nur wenig in das Angriffsspiel ihrer Mannschaft eingebunden sind. Zudem fällt auf, dass diese Spieler

tendenziell die schwächsten Passquoten aller Torspieler aufweisen, was der Grund für ihre geringe Einbindung sein könnte. Der Vergleich mit Abbildung 23 unterstützt die Charakterisierung von Kevin Trapp, Lukas Hradecky, Ralf Fährmann (Jiri Pavelnka und Alexander Schwolow) als Torspieler 2.0 (vgl. 6.1.4). Die große Anzahl von Spielern im Medianbereich der Live-Pässe bestätigt, dass der dominierende Torspieler typ derzeit der Torspieler 3.0 ist. Wenn es in der Bundesliga aktuell eine Mannschaft gibt, die hinsichtlich der Einbindung ins Angriffsspiel nahe an das Idealbild vom Torspieler 4.0 herankommt, scheint dies Borussia Mönchengladbach unter Daniel Farke zu sein. Sowohl Jonas Omlin als auch Yann Sommer (in seiner Zeit bei Gladbach) bewegen sich in Bezug auf Live-Pässe „in einer anderen Liga“ und selbst Tobias Sippel, der eigentlich als klassischer Linientorhüter aus der Ehrmann-Schule bekannt ist, wird durchschnittlich oft in das Angriffsspiel eingebunden. Einschränkend ist natürlich zu bemerken, dass alle drei (extreme) Defizite im Mitspielverhalten bei gegnerischem Ballbesitz aufweisen, was diese Einschätzung relativiert.

Die Abbildung 34 enthält Informationen über die Live Touches (Ballkontakte) und die Ballannahmen des Torspielers im Angriffsspiel seiner Mannschaft. Diese Daten ergänzen die Einschätzungen hinsichtlich der aktiven Einbindung des Torspielers in das Spielgeschehen.

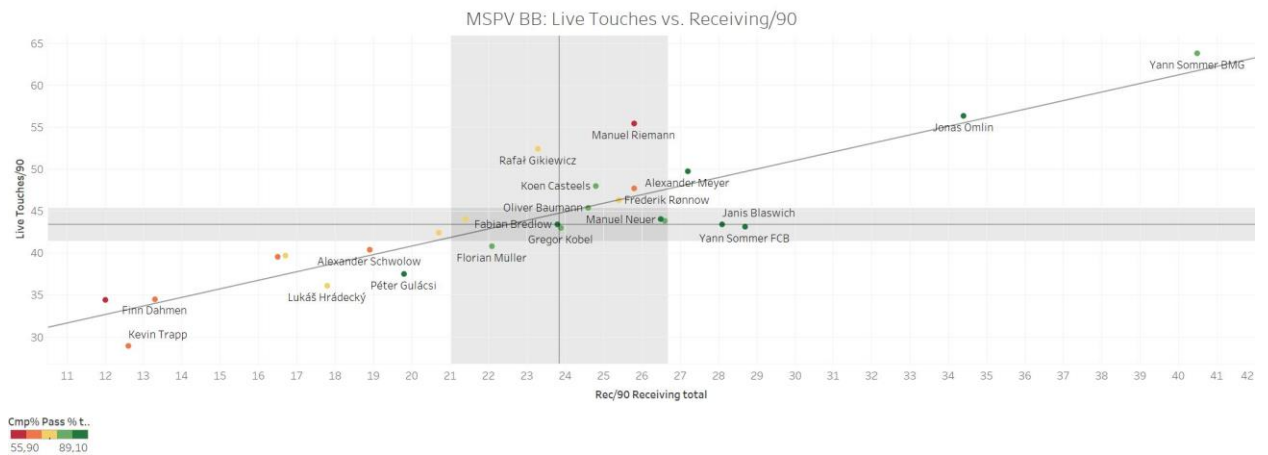


Abbildung 34: Live Touches vs. Receiving/90 Quelle: eigene Darstellung

Es besteht die Möglichkeit, die Ballkontakte des Torspielers in verschiedenen Zonen des Spielfelds zu differenzieren, wie es in Abbildung 35 dargestellt ist. Dadurch können zusätzliche Erkenntnisse über die Positionierung des Torspielers im eigenen Angriffsspiel gewonnen werden.

Es erfolgt eine Unterscheidung der Ballkontakte des Torspielers in seinem eigenen Strafraum, im ersten und im mittleren Drittel des Spielfelds.

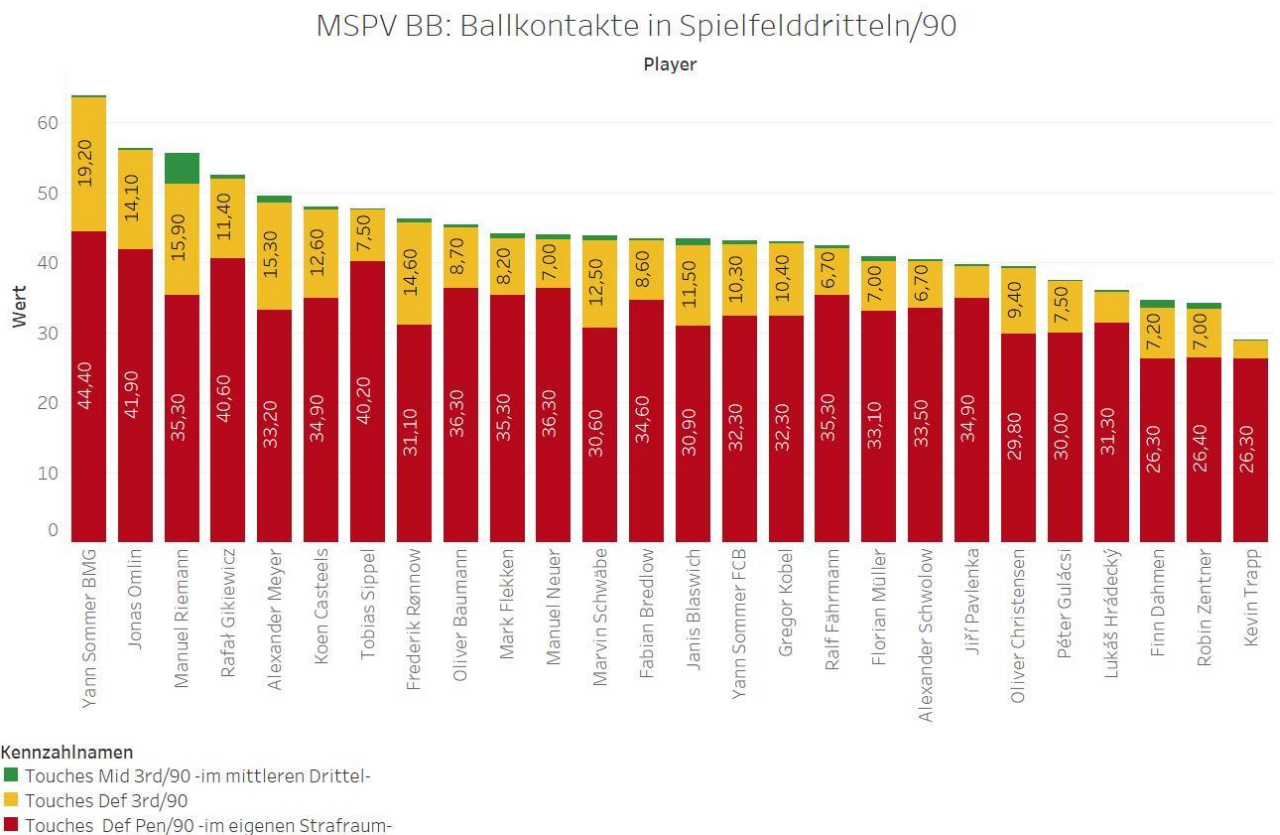


Abbildung 35: Ballkontakte nach Spielfelddritteln/90 Quelle: eigene Darstellung

Anschließend erfolgt eine detaillierte Betrachtung der Handlungen der Torspieler, wenn sie in Ballbesitz sind. Hierbei wird die „Total Distanz Carries“, also die Strecke, die der Spieler mit dem Ball am Fuß in jede Richtung zurücklegt, in Relation zur Progressiven Distanz gesetzt, wie es in Abbildung 36 dargestellt ist. Ziel ist es, jene Torspieler zu identifizieren, die mit dem Ball andribbeln, um gegnerische Stürmer zu binden und dadurch situationsbedingte Überzahlsituationen auf anderen Ebenen des Spielfelds zu schaffen. Zudem wurde die Anzahl der progressiven Pässe durch farbliche Markierung hervorgehoben.

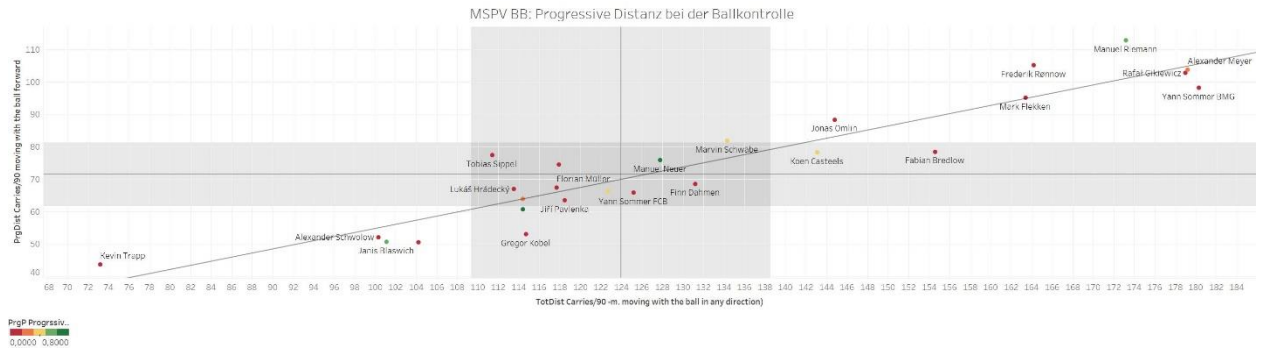


Abbildung 36: Progressive Distanz bei der Ballkontrolle Quelle: eigene Darstellung

Zum Abschluss erfolgt eine Analyse der Pässe bezüglich ihrer Progressivität, um jene Torspieler zu identifizieren, die das Spiel nicht nur durch Querpässe verlagern, sondern gezielt Vertikalpässe spielen, um gegnerische Linien zu durchbrechen oder zu überspielen (vgl. Abb. 37). Die Passquote ist farblich hervorgehoben.

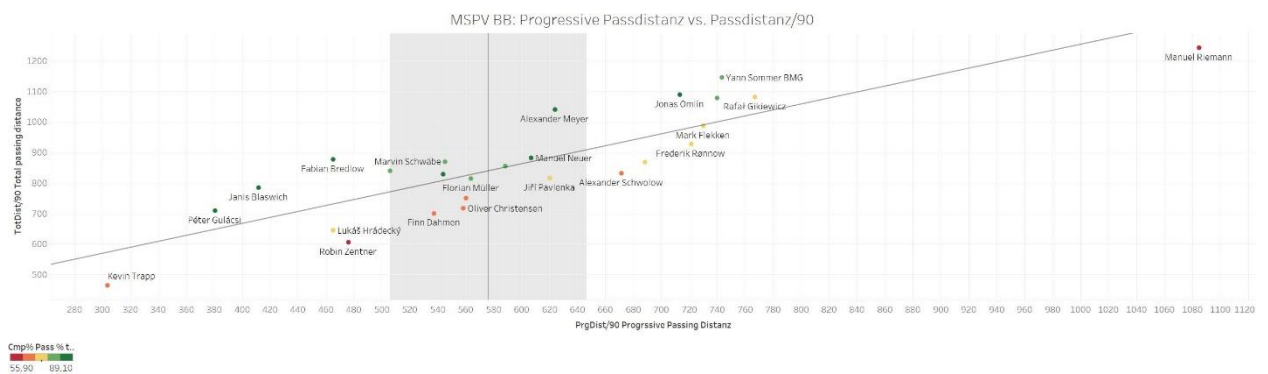


Abbildung 37: Progressive Passdistanz vs. totale Passdistanz/90 Quelle: eigene Darstellung

## 6.2 Informationsgehalt der Ergebnisse

Zusammenfassend lässt sich feststellen, dass die Ergebnisse trotz ausschließlicher Verwendung frei zugänglicher Datensätze allen wissenschaftlichen Gütekriterien entsprechen. Um die Validität, d.h. den Grad der Genauigkeit, mit der das jeweilige Ereignis oder die Metrik genau das misst, was für den Scouting-Prozess relevant ist, weiter zu erhöhen, ist eine spielkontextbezogene Adjustierung erforderlich. Dies wurde insbesondere im Bereich der Zielverteidigung erfolgreich umgesetzt. Im Bereich des Mitspielverhaltens im eigenen Ballbesitz war dies jedoch aufgrund fehlender Mannschaftsdaten nicht möglich (vgl. 5.2). Daher könnte dieser Analysebereich trotz der vorhandenen Detailtiefe und der Kontextbezüge innerhalb der Kategorie leichte Verzerrungen der tatsächlichen Leistungsfähigkeit aufweisen. Das Gleiche gilt für den Bereich des Mitspielverhaltens bei gegnerischem Ballbesitz, der derzeit unabhängig vom Anbieter noch nicht die gewünschte Detailtiefe aufweist.

Trotz dieser Einschränkungen und der Komplexität des torspieler-spezifischen Anforderungsprofils liefern die Ergebnisse einen objektiven Überblick zur tatsächlichen Leistungsfähigkeit und das individuelle Profil jedes einzelnen Spielers. Aus diesem Grund eignet sich der datenanalytische Ansatz hervorragend, um eine Vorauswahl (Short-List) potenzieller Neuzugänge zu erstellen, die dann durch Live-/Video-Scouting verfeinert werden kann. Obwohl keine detaillierte qualitative Analyse der Bundesliga-Torspieler durchgeführt wurde, sondern die Daten nur mit subjektiven Experten-Eindrücken aus Spielbeobachtungen verglichen werden, lassen sich Teilelemente identifizieren, die in der qualitativen Analyse möglicherweise erst später aufgefallen wären. Als Beispiel können hier die Defizite von Yann Sommer im Mitspielverhalten bei gegnerischem Ballbesitz oder von Kevin Trapp im Mitspielverhalten allgemein genannt werden. Diese auf Daten-Basis identifizierten Teilelemente, unabhängig davon, ob sie positiver oder negativer Natur sind, denen in der rein qualitativen Analyse möglicherweise nur geringe Bedeutung beigemessen worden wäre, müssen durch die Kombination der Methoden in der qualitativen Analyse verifiziert oder falsifiziert werden. Anschließend muss beurteilt werden, ob diese Faktoren ein Ausschlusskriterium oder die entscheidende Fähigkeit sind, warum ein bestimmter Torspieler gegenüber einem anderen präferiert wird.

Eine Diskussion über die Objektivität der Clusteranalyse ist durchaus angebracht, da sie auf selbstberechneten KPIs basiert. Die im Rahmen dieser Arbeit erstellten KPIs sind aufgrund der Gewichtung bestimmter Elemente innerhalb der Kategorien zwangsläufig subjektiv. Dies führt dazu, dass Verhaltensweisen, die der Autor bevorzugt, überbewertet werden und solche, die als eher kontraproduktiv erachtet werden, unterbewertet werden können. Daher ist es möglich, dass Experten mit unterschiedlichen Philosophien unterschiedliche Meinungen zur Gewichtung einzelner Elemente haben, ohne dass es eine eindeutig richtige oder falsche Antwort gibt. In der praktischen Anwendung des Scoutings im Vereins-/Verbandskontext ist es vor allem wichtig, dass die Elemente so gewichtet werden, wie es im vereinsspezifischen Positionsprofil notwendig ist. Dies bedeutet, dass sich aus der Spielphilosophie bestimmte Situationen ergeben, die der jeweilige Torspieler häufiger, besser oder anders lösen muss als der Durchschnitt, sowie Situationen oder Teile des Anforderungsprofils, die vernachlässigt werden können, da sie weniger häufig oder ohne besonderen Qualitätsanspruch auftreten.

Die Identifikation von Indikatoren für vereinsspezifische Anforderungen im Mitspielverhalten im eigenen Ballbesitz kann unter Berücksichtigung des Ortes (Position auf dem Spielfeld), der Drucksituation und der Zielspieler bzw. Zielräume im Spielaufbau erfolgen. Daraus ergibt sich die Frage nach der Notwendigkeit von Vielseitigkeit in den Angriffstechniken des Torspielers im Hinblick auf die Spieleröffnung, wobei bestimmte Waffen wie der Abschlag, Abwurf oder der Flugball eine Rolle spielen können, wenn es um das schnelle Umschalten nach Ballgewinn geht. Im Mitspielverhalten bei gegnerischem Ballbesitz ist die Höhe der Kette in den verschiedenen

Spielphasen von entscheidender Bedeutung. Wenn Mannschaften im eigenen Ballbesitz tendenziell kompakt angreifen und eine hohe, numerisch ausgeglichene Restverteidigung (z.B. 1vs1 oder 2vs2) nutzen, um bei einem möglichen Ballverlust im vorderen Drittel durch Gleich- oder Überzahl sofort ins Gegenpressing zu gehen und/oder im Ballgewinnspiel vorwiegend das Angriffspressing zu nutzen, werden für den Gegner hinter der Kette größere Zielräume entstehen, die der jeweilige Torspieler beherrschen sollte. Hingegen ist diese Sweeper-Fähigkeit bei Mannschaften, die tendenziell tiefer stehen und nach Ballverlust sofort zurückfallen, von untergeordneter Bedeutung. Zudem spielt die Frage, wie oft und welche Art von Flanken/Querpassen (bewusst) zugelassen werden, eine entscheidende Rolle für das vereinsspezifische Anforderungsprofil im Bereich des Mitspielverhaltens bei gegnerischem Ballbesitz. Im Bereich der Zielverteidigung empfiehlt es sich zu untersuchen, aus welchen Spielsituationen die meisten und/oder spielentscheidenden torkritischen Situationen des Gegners entstehen oder auch bewusst zugelassen werden sollen. Insbesondere in Spitzenmannschaften spielen die Torspieler vs.1-Situationen, obwohl sie nur einen geringen Anteil der Gesamtktionen ausmachen, eine spielentscheidende Rolle. Wenn man weiß, dass die eigene Mannschaft überdurchschnittlich oft mit solchen Situationen konfrontiert ist, sollte man auch einen Torspieler auswählen, der in diesem Bereich besondere Stärken oder zumindest keine Schwächen aufweist.

In einer umfassenden Bewertung auf der Grundlage dieser Datenanalyse lässt sich feststellen, dass die Torspieler in der Bundesliga im Angriffsspiel noch keine zentrale Rolle einnehmen. Mit Ausnahme von Borussia Mönchengladbach scheint keine Mannschaft dieses Potenzial konsequent nutzen zu wollen, um numerische Überzahlsituationen zu schaffen. Selbst Bayern München, wo Manuel Neuer und Yann Sommer alle erforderlichen Fähigkeiten mitbringen würden um das Angriffsspiel des Teams auf ein höheres Niveau zu bringen, hat diesen Ansatz unter Nagelsmann in Situationen, in denen es erforderlich gewesen wäre, viel zu selten genutzt. Hoffenheim, die in der Vergangenheit bereits nahe am Idealbild waren, tendiert in dieser Saison auch eher zum Torspieler 3.0. Der Torspielerotyp 4.0 (vgl.4.1) scheint deshalb in der aktuellen Bundesliga-Saison quasi nicht zu existieren.

Dies liegt vor allem an der Charakteristik der Liga, die mehr von schnellem Umschaltspiel als von dominierendem Ballbesitzspiel geprägt ist. Aus Sicht des Torspielers würde diese These im Umkehrschluss bedeuten, dass insbesondere das Mitspielverhalten bei gegnerischem Ballbesitz bzw. direkt im Moment des Umschaltens nach Ballverlust von herausragender Bedeutung sein sollte. Dies kann anhand der Daten jedoch nicht bestätigt werden. Im Gegenteil, erstaunlich viele der aktuellen Bundesliga-Torspieler weisen eklatante Defizite in diesem Bereich auf, sodass einige sogar als Torspieler 2.0 charakterisiert werden müssten. Dies ist ein Torspielerotyp, der vor 15 bis 20 Jahren als "State of the Art" galt. Auch wenn die analysierten Bundesliga-Spieler sicherlich nicht als vollständig repräsentativ für das gesamte Ausbildungssystem angesehen

Janik Backhaus



werden können (auch weil viele Ausländer dabei sind, die nicht im deutschen System ausgebildet wurden), zeigt diese Beobachtung die Defizite im Training und in der Trainerausbildung auf, die im großen Ganzen dafür verantwortlich sind, dass sich Deutschland Sorgen machen muss, ob auch in Zukunft Weltklasse-Torspieler entwickelt werden. Denn Weltklasse entsteht in der Regel, unabhängig von der Position, auf Basis umfassender Fähigkeiten, die auf einem sehr hohen Niveau beherrscht werden, sowie einzelner "Waffen", die die Spieler besonders machen. Das bedeutet, wenn bestimmte Teilbereiche im Training und in der Trainerausbildung vernachlässigt werden, kann unabhängig davon, wie gut in den anderen Teilbereichen gearbeitet wird, keine Weltklasse entstehen.

Unabhängig davon lässt sich im ligaübergreifenden Vergleich feststellen, dass der Torspieler Typ 4.0 noch auf seinen endgültigen, flächendeckenden Durchbruch wartet. Es gibt prominente Beispiele wie den FC Barcelona (ter Stegen) oder Manchester City (Ederson) sowie Vereine in der zweiten Liga wie den HSV (Heuer Fernandes) und Magdeburg (Reimann), die den Torspieler 4.0 bereits fest in ihr Spielkonzept integriert haben, aber insgesamt doch eher wenige. Letztlich hängt die weitestgehend spielstands- und spielphasenunabhängige Art der Integration des Torspielers in den Spielaufbau von der Risikobereitschaft und dem Mut des Cheftrainers ab. Es ist zu beachten, dass eine stärkere Integration des Torspielers im Falle eines Ballverlusts eine potenziell höhere Gefahr eines einfachen Gegentors mit sich bringt, die im Verhältnis zum Mehrwert des Überzahlspielers im Angriffsspiel abgewogen werden muss. Es besteht jedoch Einigkeit darüber, dass es Spielsituationen und Spielphasen gibt, in denen der Torspieler nicht ausschließlich als letzte Option betrachtet werden sollte, um den Ballbesitz zu sichern, sondern eine zentrale Rolle einnehmen kann, um aus dem Ballbesitzspiel ein effektives, vertikales Angriffsspiel zu initiieren. Die Beantwortung der Frage, ob, in welchem Ausmaß und bei welchen Spielern/Vereinen dies bereits in der Bundesliga beobachtet werden kann, ist im Rahmen dieser übergreifenden Datenanalyse nicht möglich. Um eine abschließende Feststellung zu treffen, wäre eine Analyse auf der Ebene einzelner Spiele, bzw. sogar Spielsituationen erforderlich.

Wenn das Ziel der vorliegenden Arbeit darin bestanden hätte, den besten und komplettesten Torspieler der Bundesliga zu identifizieren, wäre Manuel Neuer nach wie vor der herausragende Kandidat, da er auf Basis der Datenanalyse als einziger in nahezu allen Teilbereichen in den Top-3 positioniert ist. Möglicherweise könnte diskutiert werden, ob Yann Sommer aufgrund seiner vermeintlich größeren Stärken im Bereich der Zielverteidigung und des Mitspielverhaltens im eigenen Ballbesitz im Vergleich zu Neuers deutlichen Vorteilen im Mitspielverhalten im gegnerischen Ballbesitz eine höhere Wertung verdient. Aufgrund der vorliegenden Daten gibt es darüber hinaus jedoch keinen großen Spielraum für Diskussionen. Peter Gulasci, Gregor Kobl und Frederik Rönnow würden auf den Plätzen 3-5 folgen, da sie herausragende Fähigkeiten in der Zielverteidigung haben, aber in beiden Komponenten des Mitspielverhaltens eher im

Janik Backhaus

Mittelmaß anzusiedeln sind. Ebenfalls auf Datenbasis unzweifelhaft ist, dass sich Oliver Christensen, Alexander Schwolow, Florian Müller, Jiri Pavlenka und Fabian Bredlow am unteren Ende eines potenziellen Rankings befinden, während Oliver Baumann und Marvin Schwäbe überraschenderweise nur knapp darüber liegen. Die Spieler dazwischen liegen tatsächlich alle relativ nah beieinander, und es könnte darüber diskutiert werden, welche Fähigkeit wichtiger ist.

Wenn das Ziel darin bestanden hätte, einen Torspieler zu identifizieren, der sich unabhängig von den spezifischen Anforderungen eines Vereins für höhere Aufgaben empfiehlt, dann wäre Finn Dahmen zweifellos die geeignete Wahl. In Bezug auf die Zielverteidigung und das Mitspielverhalten bei gegnerischem Ballbesitz befindet er sich durchweg im besten Quadranten oder im oberen Drittel, obwohl er kein einziges Spiel zu Null gespielt hat. Die vermeintlichen Mängel im eigenen Ballbesitz scheinen mit der Spielphilosophie seines Teams zusammenzuhängen, da sein Konkurrent Robin Zentner ähnliche oder sogar schlechtere Werte aufweist. Dahmen, dessen Vertrag ausläuft, konnte sich aus Gründen, die auf Datenbasis nicht erklärbar sind, bei Mainz nicht dauerhaft als Nummer Eins etablieren. In den sechs Spielen, in denen er eingesetzt wurde, hat er jedoch konstant gezeigt, dass er Mannschaften verstärken kann, die sich auf einem ähnlichen oder sogar höheren Niveau als Mainz 05 (gehobenes Mittelfeld im Kampf um die internationalen Plätze) befinden. Es ist anzumerken, dass Finn Dahmen als U21-Europameister sicher kein Geheimtipp mehr ist und längst einen Stammplatz in der Bundesliga haben sollte. Diant Ramaj und Jan Jakob Olschowsky hingegen können als Geheimtipps betrachtet werden, da sie nur in einem oder zwei Spielen eingesetzt wurden und aufgrund des 20%-Spielzeit-Filters in den Analysen ab 6.1.3 nicht mehr berücksichtigt wurden. Trotzdem haben sie in diesen Spielen mehr als nur angedeutet, dass sie das Potenzial haben, langfristig in der Bundesliga zu spielen. Die entsprechenden quantitativen Daten, die diese These untermauern würden, würden den Rahmen dieser Arbeit sprengen, können jedoch bei Bedarf zur Verfügung gestellt werden.

## **7 Ausblick**

Im letzten Kapitel der Arbeit werden verschiedene Möglichkeiten zur Verbesserung des vorgestellten Ansatzes aufgezeigt und ein Ausblick auf die mögliche zukünftige Entwicklung im Bereich der Datenanalyse gegeben.

### **7.1 Mögliche Erweiterung des eigenen Ansatzes**

Eine potenzielle Erweiterung des vorgestellten Ansatzes hängt letztendlich von der Verfügbarkeit, Qualität und Validität der zur Verfügung stehenden Daten ab. Wie bereits in den Abschnitten 5.2 und 6.2 angedeutet wurde, hätten die Daten zum Mitspielerverhalten wesentlich aussagekräftiger sein können, wenn klar zuordenbare Daten auf Mannschaftsebene zur Verfügung gestanden

hätten. Es wurde zwar der Versuch unternommen, aber letztendlich lässt sich aus dem Gesamtwert einer Mannschaft über 24 Spieltage nicht der exakte Wert für 3,5 oder 10 Spiele extrahieren. Die Annahme, dass die Durchschnittswerte zur Adjustierung verwendet werden können, führt dazu, dass die Werte der Spieler, die viele oder sogar alle Spiele absolviert haben valider erscheinen, aber gleichzeitig die Werte aller anderen Spieler verzerrt werden.

Die Verwendung eines umfassenden Datensatzes von führenden Anbietern wie Opta, StatsBomb oder Sports-Tec-Solution hätte zweifellos die Detailtiefe erhöht. Es ist jedoch fraglich, ob dies zu grundlegend unterschiedlichen Ergebnissen oder Schlussfolgerungen geführt hätte. Um eine präzise Antwort auf die potenziellen Auswirkungen dieser Datenquellen auf den vorgestellten Ansatz zu geben, sind weitere Forschung und vergleichende Analysen erforderlich. Die Frage nach der größten Validität aller verfügbaren Datensätze wird durch die Verwendung unterschiedlicher Datenquellen jedoch weiterhin nicht eindeutig beantwortet werden können. Es ist zu erwarten, dass es aufgrund von Unterschieden in der Anzahl der verfügbaren Daten und Metriken, sowie deren Definitionen und Berechnungsmethoden, zu Variationen kommt (vgl. 4.2). Diese Unterschiede können dazu führen, dass bestimmte Verhaltensmuster der Torspieler besser, schlechter oder nach wie vor nicht ableitbar oder quantifizierbar sind.

Unabhängig vom Anbieter gibt es im Bereich der Datenerfassung jedoch noch zahlreiche ungenutzte Potenziale, die es ermöglichen würden, eine detailliertere und objektivere Beurteilung der Leistungsfähigkeit von Torspielern vorzunehmen. Insbesondere werden in den nachfolgend skizzierten Ansätzen Tracking-/Positions- und Eventdaten kombiniert, die bereits mit den heutigen technischen Möglichkeiten der optischen Objekterkennung weitgehend automatisiert erfasst werden könnten oder sogar bereits für Feldspieler erfasst werden.

### **7.1.1 Optimierungspotenziale Spieldatenerfassung - Zielverteidigung**

Eine bedeutende Innovation im Bereich der Zielverteidigung könnte durch die Nutzung von Daten zum Stellungsspiel des Torspielers erzielt werden. Experten stimmen überein, dass dies die entscheidende Fähigkeit zur Verhinderung von Gegentoren darstellt. Der Ausgangspunkt hierfür ist die Berechnung der virtuellen Trefferfläche, die zum Zeitpunkt des Torschusses (T) entsteht. Diese wird durch den Schnittpunkt der Körperachse des Torspielers (variable X-Y-Koordinate) mit den vier virtuellen Linien bestimmt, die den Ball (variable X-Y-Koordinaten) mit dem oberen und unteren Ende der Torpfosten (fixierte X-Y-Z-Koordinaten) verbinden. Die Angabe der Trefferfläche als absolute Zahl oder als Verkleinerung im Verhältnis zur ursprünglichen Torfläche ist nicht entscheidend. Um die Verzerrung durch die zwangsläufige Abhängigkeit von der Torschussposition zu minimieren, kann zwischen zentralen und seitlichen Aktionen unterschieden werden. Es gilt die Regel, dass je kleiner die Trefferfläche ist, desto besser die Chance, einen Treffer zu verhindern, sofern der Torspieler aktionsbereit ist.

Wir kommen damit zur Ableitung der Aktionsbereitschaft als zweitem wichtigen KPI. Die Frage ist, ob der Torspieler zum Zeitpunkt T mit beiden Füßen in der Grundposition aktionsbereit ist oder nicht. Diese Frage kann mit "Ja" oder "Nein" beantwortet werden. Wenn die Antwort "Ja" lautet, ist es wichtig zu ermitteln, wie viel früher (T-X) der Torspieler bereit war. Wenn die Antwort "Nein" lautet, ist es wichtig zu ermitteln, wie viel später (T-X) der Torspieler bereit war. Grundsätzlich gilt, dass eine höhere zeitliche Präzision, welche durch einen geringen Abstand zum Zeitpunkt T bzw. 0 ausgedrückt wird, als positiv angesehen wird. Zudem ist es besser, eher etwas zu früh als zu spät aktionsbereit zu sein. Darüber hinaus lässt sich durch die Bestimmung des Startzeitpunkts der abwehrenden Aktion über 3-Modeling die Entscheidungsfindung im magischen Dreieck zwischen Reagieren, Antizipieren und Spekulieren, als dritter entscheidender KPI, ableiten.

Im Zusammenhang mit dem Stellungsspiel des Torspielers soll hier der vierte potenzielle KPI, die mathematische Bestimmung der optimalen Position, kurz erwähnt werden. Diese ist aufgrund ihrer scheinbaren Komplexität in der Praxis umstritten, da sich die Philosophien der Trainingslehre unterscheiden und die mathematisch-physikalische Logik noch nicht allgemein akzeptiert ist. Unter Bezugnahme auf Karl-Heinz Rummenigge kann man feststellen, dass seine These "Fußball ist keine Mathematik" in diesem Kontext leider weit verbreitet scheint. Es ist jedoch zu bemerken, dass selbst ein grundlegendes Verständnis von Mathematik im Situationskontext einen signifikanten Mehrwert bieten würde. Die optimale Position in Bezug auf die Spielfeldbreite (X-Koordinate) kann relativ einfach über die virtuelle Verbindung zwischen Ball und Tormitte ermittelt werden, da der kürzeste Weg zu beiden Seiten (Pfosten) immer in der Mitte beginnt, oder mathematisch ausgedrückt auf der Winkelhalbierenden der Trefferfläche liegt. D.h. die optimale Position befindet sich im Allgemeinen auf der gedachten Linie zwischen Ball und Tormitte bzw. Winkelhalbierende der Trefferfläche, es sei denn, ein Teil der Trefferfläche wird durch einen oder mehrere Mitspieler abgedeckt. In diesem Fall, in dem Torspieler und Mitspieler kooperieren, um Torschüsse zu blocken, muss eine neue virtuelle Trefferfläche durch Berücksichtigung der X-Y-Koordinaten der Mitspieler (die einen oder mehrere Pfosten „ersetzen“) berechnet werden. Die optimale Position des Torspielers verschiebt sich dann von der Verbindung zur Tormitte auf die Winkelhalbierende der virtuellen Trefferfläche (vgl. Backhaus et. al., 2021, S.32ff.).

Für die Y-Koordinate gilt nach dem Prinzip des Ballangriffs, dass die Trefferfläche so weit wie möglich zu verkleinern ist, d.h. tendenziell eine höhere Ausgangsposition eingenommen werden sollte, um den gleichen Ball früher und somit körpernäher angreifen zu können. Dadurch können fehlende Sprungkraft, Körpergröße oder eine schlechtere Position in der Breite zumindest in Teilen ausgeglichen werden, weil der Weg zum Ball einfach deutlich kürzer ist. Ein weit verbreiteter Irrglaube besteht jedoch in der Annahme, dass sich der Torspieler durch eine tiefere

Janik Backhaus

Position auf dem Spielfeld Vorteile hinsichtlich seiner Reaktionszeit verschaffen würde. Wenn man die Geschwindigkeit des Balls berücksichtigt, stellt man schnell fest, dass es nur um einen Zeitgewinn von wenigen Hundertstel, wenn nicht gar Tausendstel Sekunden geht. Angesichts der menschlichen Reaktionszeit von 0,2 bis 0,3 Sekunden hat dieser Zeitgewinn kaum einen Mehrwert. Zudem bleibt in diesem Ansatz der Nachteil der vergrößerten Trefferfläche durch die tiefere Position unberücksichtigt, der sich in einer anderen Dimension des Raum-Zeit-Kontinuums und je nach Schusswinkel in einem viel größeren Wertebereich bewegt als der Zeitgewinn. Derzeit gibt es mit Stats Bomb nur einen Anbieter, der den Mehrwert dieses KPIs erkannt hat und den Mut hatte ein erstes berechnendes Modell zu entwickeln, welches jedoch leider auch von diesem Irrglauben beeinflusst ist (vgl. 4.2). Dieses Modell kann weiterentwickelt werden, um die berechnete Idealposition auch auf der Y-Achse näher an die praxisrelevante und erreichbare Idealposition heranzuführen. Unter darüberhinausgehender Berücksichtigung der Ausnahmesituation der "curved Balls", d.h. der Schüsse, die aufgrund von Drehungen oder Absenkungen in ihrer Flugbahn von einem direkten Weg zum Tor abweichen, wird die Schwierigkeit der Erstellung eines Modells mit größtmöglicher Validität nochmals deutlich. Es bleibt jedoch unklar, ob es bereits Algorithmen gibt, die die Ausnahmen von der Regel im Kontext der Komplexität der Situation automatisiert erkennen und die in der Folge notwendigen Berechnungen durchführen können. Eine solche Entwicklung würde zweifellos einen Quantensprung in der datenbasierten Analyse des Torspielers bedeuten.

In der Praxis ist es neben dem Stellungsspiel natürlich von entscheidender Bedeutung, wie erfolgreich der Torspieler im Abwehren von Schüssen ist. Eine sinnvolle Ergänzung zu den vorhandenen Eventdaten, wenigen Advanced Metrics und insbesondere zum Post-Shot-xG-Wert wäre ein "Expected Save"-Wert, welcher die Wahrscheinlichkeit des Torspielers angibt, den jeweiligen Torschuss zu halten. Dies erfordert im Wesentlichen lediglich eine Anpassung der Metrikperspektive.

### **7.1.2 Optimierungspotenziale Spieldatenerfassung - Mitspielverhalten im eigenen Ballbesitz**

Im Bereich des Mitspielverhaltens im eigenen Ballbesitz besteht ein Bedarf an einer gezielten Erfassung von Event-Attributen und Advanced Metrics, welche bereits für Feldspieler erhoben werden. Insbesondere bestimmte Kennzahlen aus dem Packing-Kontext des Impact-Portfolios spielen hierbei eine entscheidende Rolle. Diese Kennzahlen sind in Teilen unter veränderten Begriffsbezeichnungen oder leicht abgewandelt auch von anderen Anbietern verfügbar. Bedauerlicherweise beschränkt sich das Angebot meist ausschließlich auf die Erfassung von Feldspielern, während artgleiche Aktionen von Torspielern entweder gar nicht oder nur unzureichend erfasst werden.

Zur Messung der Effektivität des Torspielers im Spielaufbau bieten sich verschiedene Metriken an: Insbesondere die Anzahl der überspielten Gegen- und Mitspieler (jeweils getrennt erfasst als  $\varnothing/90$  &  $\varnothing/Pass$ ) wäre ein entscheidender Indikator. Ergänzend hierzu kann die Differenz zwischen den Pässen erfasst werden, mit denen ein zuvor eigentlich überspielter Gegenspieler wieder ins Spiel gebracht wurde (Gegner rein) und den Pässen, mit denen Gegner überspielt bzw. aus dem Spiel genommen wurden (Gegner raus). Diese Metrik kann sowohl als Differenz der Passanzahl als auch als Differenz der Gegenspieler dargestellt werden und wiederum als Durchschnitt pro 90 Minuten oder Pass erfasst werden. Ein weiteres relevantes Passattribut sind Querpässe, bei denen weder Gegenspieler aus dem Spiel genommen noch wieder ins Spiel gebracht wurden. Eine prozentuale Aufteilung dieser drei Passattribute würde zudem einen enormen Mehrwert bieten. Aus diesen drei Attributen lässt sich zudem ein aussagekräftiges "Expected Pass"-Modell entwickeln, welches die Qualität der Entscheidungsfindung und die Risikobereitschaft des Torspielers widerspiegelt. Zusätzlich wären aus Torspielersicht die Advanced Metrics "Expected Threat (xT)" bzw. "xG-Buildup" interessant, da sie den Beitrag des Torspielers in der frühen Phase des Angriffsspiels würdigen. Der Beitrag des Torspielers lässt sich nicht allein durch Assists, expected Assists oder pre-Assists bewerten, da der Pass des Torspielers in der Regel nicht der letzte oder vorletzte vor einem Torabschluss ist.

Um qualitative Unterschiede zwischen Ballbesitzphasen des Torspielers zu ermitteln, könnte die Betrachtung der Performance under Pressure, also der Leistung unter Druck, herangezogen werden. Dazu müsste die Distanz zwischen der Position des ballbesitzenden Torspielers und dem nächsten Gegenspieler zum Zeitpunkt des Passes ermittelt werden. Diese Distanz sollte in verschiedene Druckzonen unterteilt werden, wie beispielsweise  $< 5$  Meter = hoher Gegnerdruck,  $5-10$  Meter = mittlerer Gegnerdruck und  $>10$  Meter = niedriger Gegnerdruck. Eine anschließende Analyse der Veränderung bevorzugter Passmuster (Gegner raus, Gegner rein, Quer – oder schlicht kurz, mittel, lang) und der Passquote wäre von großem Interesse. Es wäre ebenso möglich, die Drucksituation des angespielten Mitspielers auf ähnliche Weise zu quantifizieren. Schließlich könnte das Verhältnis von "Raus aus Druck" vs. "Rein in Druck" als hervorragender KPI für die Qualität der Spielfortsetzung durch den Torspieler herangezogen werden.

Zusätzlich von Bedeutung sind die kritischen Ballverluste des Torspielers während der Spieleröffnung. Obwohl diese nicht sehr häufig vorkommen, wie aus den Passquoten für kurze und mittlere Distanzen hervorgeht (vgl. 6.1.4.), können sie in der Regel spielentscheidend sein. Hierbei ist möglicherweise eine geringfügige Anpassung spezifisch für den Torspieler erforderlich, da bei einem Ballverlust oder Fehlpass des Torspielers im ersten oder zweiten Drittel des Spielfelds in der Regel nur noch sehr wenige Spieler hinter dem Ball sind. Folglich sollte die Anzahl der eigenen Spieler, die durch den Ballverlust aus dem Spiel genommen wurden, möglicherweise von derzeit 4 auf 6 oder 7 erhöht werden. Es wäre auch denkbar, jene

Ballverluste als kritisch zu definieren, bei denen eine Gleichzahl- (einschließlich des Torspielers) oder sogar Unterzahl-Situation hinter dem Ball entsteht.

Um eine genauere Beurteilung des Freilaufverhaltens des Torspielers zu ermöglichen, wäre seine durchschnittliche Distanz zur eigenen Abwehrkette und der gegnerischen Angriffsreihe von entscheidender Bedeutung. Dies würde einen deutlich spezifischeren Kontext bieten als zum Beispiel die Analyse der Ballkontakte in bestimmten Zonen des Spielfelds. Hierbei könnten auch die Einbeziehung des Torspielers in reaktische Aufstellungen im Ballbesitz (einzelspielbezogen) und die Erstellung spielphasenbezogener Heatmaps oder Actionmaps (einzelspielbezogen und spielübergreifend) eine Bereicherung sein. Da das Freilaufverhalten/Stellungsspiel des Torspielers im eigenen Ballbesitz gleichzeitig eine Doppelfunktion im Kontext eines möglichen Ballverlusts und des schnellen Umschaltens des Gegners erfüllt (vgl. 4.2), wären diese Metriken auch für das Stellungsspiel bei gegnerischem Ballbesitz von hoher Bedeutung.

### **7.1.3 Optimierungspotenziale Spieldatenerfassung - Mitspielerverhalten bei gegnerischem Ballbesitz**

Zusätzlich zu den in Abschnitt 7.1.2 erwähnten Heatmaps, Actionmaps und Distanzen zu Mit- und Gegenspielern, die als Indikatoren für das Mitspielerverhalten bei gegnerischen Ballbesitzes dienen, ist die Art der Balleroberung und die anschließende Spielfortsetzung durch den Torspieler von entscheidender Bedeutung. Es ist wichtig zu betonen, dass letztgenannten Kriterien auch der Spieleröffnung (d.h. eigentlich dem Mitspielerverhalten im eigenen Ballbesitz) zugeordnet werden können. Im Gesamtkontext der Optimierungspotenziale in der Datenerfassung in den einzelnen Spielphasen erscheint die Einordnung in diese Kategorie angemessener. Dies liegt daran, dass es sich um den unmittelbaren Umschaltmoment nach dem Ballgewinn durch den Torspieler handelt, der durch sein Mitspielerverhalten bei gegnerischem Ballbesitzes ausgelöst wurde.

Ein sinnvoller erster Schritt zur Verbesserung der Datenerfassung wäre eine mindestens zweiteilige Unterscheidung der Outside Penalty-Area-Actions. Hierbei sollten tatsächliche Balleroberungen und Aktionen, bei denen der Ballbesitz unmittelbar wieder zum Gegner übergeht (z.B. Clearances oder Befreiungsschläge) unterschieden werden. Eine zeitliche Erfassung der Anzahl kritischer Ballverluste, die unmittelbar nach Balleroberung entstehen, würde einen noch größeren Mehrwert bieten. Zusätzlich wären Balleroberungsversuche interessant, bei denen der Torspieler versucht hat, einen Ball abzulaufen, diesen jedoch verfehlt oder nicht erreicht hat. Unabhängig von der Art der Spielfortsetzung wäre die Anzahl der Mitspieler, die vor der Balleroberung überspielt waren und durch den Ballgewinn oder die klärende Aktion des Torspielers zurück ins Spiel geholt wurden, von großer Bedeutung (Impact Metrik: Balleroberung Mitspieler rein). Gleiches gilt für die Anzahl der Gegenspieler (in der Regel Angreifer), die nach

der Aktion vor dem Ball (näher am Tor des Torspielers als der Ball) stehen und somit nicht mehr verteidigen können oder, wenn überhaupt, nur aus dem Rücken des Torspielers im sofortigen Gegenpressing mitwirken können (Impact Metrik: Gegenspieler raus).

Eine umfassende Verbindung zwischen den mitspielenden Aktionen außerhalb des Strafraums und der Strafraumbeherrschung lässt sich anhand der Anzahl aller Ballgewinne durch den Torspieler darstellen, die nicht durch gegnerische Torschüsse entstanden sind. Um die Positionen der Balleroberungen auf dem Spielfeld zu differenzieren, könnte dies mithilfe einer Actionmap erfolgen. Eine weitere Differenzierung bezüglich der Art der Flanken, wie flache, scharfe Hereingaben (Querpassverteidigung), halbhohe und hohe Flanken, wäre wünschenswert. Aus Sicht des Torspielers wären die Zielräume innerhalb des Strafraums jedoch von größerer Relevanz. Daher sollte der 5-Meterraum in drei, vier oder fünf relevante Teilbereiche (mindestens: Zone vor dem ballnahen (ersten) Pfosten, zentrale Zone, Zone hinter dem ballfernen (zweiten) Pfosten; ideal: Aufteilung der zentralen Zone nach Tordritteln/Hälften) und der restliche Teil des Strafraums in zwei oder drei Zonen (ballnah/ (zentral)/ ballfern) gegliedert werden. Es erscheint nicht sinnvoll, die bekannten Spielfeldraster zu verwenden, da der Strafraum in der Regel nur in einer vertikalen Ebene abgebildet ist und die Breite des 5 Meterraums in seiner Gesamtheit die zentrale horizontale Zone repräsentiert. Anschließend sollten die Torspieler-Aktionen (Fangen, Ablenken, Fausten) sowie die Aktivität (Balleroberungsversuch vs. Zielverteidigung auf der Linie) den Zonen zugeordnet werden. Dadurch würde ein sehr detailliertes Datenbild zur Strafraumbeherrschung eines Torspielers entstehen, welches in Anwendungsbereichen Scouting und Analyse genutzt werden kann.

Abschließend gilt es, die Fähigkeiten zur schnellen Kontereinleitung oder das Streben danach quantifizierbar zu machen. Hierbei wäre die zeitliche Differenz zwischen Ballgewinn und Spieleröffnung als erste relevante Metrik zu erfassen. Eine ideale Ergänzung dazu stellt die durchschnittliche Zeit vom Ballgewinn des Torspielers bis zur Ballkontrolle eines Mitspielers im Angriffsdrittel (inklusive zweiter Bälle), die durchaus als KPI betrachtet werden könnte, dar. Es ist jedoch zu beachten, dass die Entscheidung zur schnellen Spielfortsetzung von Faktoren wie der Staffelung des Gegners im Moment des Ballgewinns, den potenziellen Zielräumen, die von Mitspielern angelaufen werden können, dem Spielstand und der Spielphilosophie abhängt. Durch eine übergreifende Betrachtung mehrerer Spiele können jedoch jene Spieler identifiziert werden, die bewusst ihre „Waffen“ in der Spieleröffnung einsetzen, um die tendenziell immer vorhandene fehlende Ordnung des Gegners auszunutzen. Eine bloße Differenzierung nach Länge, Technik und zugehöriger Passquote der Spieleröffnung wäre hingegen von geringerem Wert aber immer noch ein Fortschritt.



## 7.2 “Wo geht die Reise hin im Datenanalysebereich”?

Aus Sicht der datenbasierten Torspieleranalyse wäre es wünschenswert, wenn führende Datenprovider die in Abschnitt 7.1 genannten Optimierungspotenziale berücksichtigen und einige der vorgeschlagenen Advanced Metrics bzw. KPIs in ihre standardisierte Datenerfassung und -verarbeitung aufnehmen würden. Dabei sollten technische Umsetzbarkeit und Validität gewährleistet sein, jedoch lässt sich der Kosten-Nutzen-Faktor nicht abschließend beurteilen. Möglicherweise bedarf es eines Pioniers auf dem Gebiet der Torspieleranalyse, sozusagen einem „Bill James der Torspieleranalyse“, der die traditionellen Analysemethoden kritisch hinterfragt, in einen spielrealistischeren Kontext setzt und innovative Ideen entwickelt, um die Spieler objektiver zu bewerten und die Gesamtleistung eines Spielers zu erfassen, anstatt nur Teilaspekte zu betrachten.

Im großen Ganzen betrachtet befindet sich die Datenanalyse im Fußball noch in den Anfängen, aber die Revolution von einer überwiegend qualitativ geprägten Analysemethodik hin zu einer immer größer werdenden Bedeutung der quantitativen Analyse hat bereits begonnen und nimmt derzeit an Fahrt auf. Eine strukturierte, qualitativ hochwertige Datenanalyse führt in der Praxis zu einer massiven Zeit- und Kostenersparnis im Vergleich zur rein qualitativen Analyse. Außerdem liefert eine Methodenkombination unabhängig vom Anwendungsbereich objektivere und somit bessere Ergebnisse. Trotzdem wird der qualitative Analyseanteil sowohl im Scouting als auch in der Spielanalyse aufgrund der Komplexität des Fußballspiels immer höher gewichtet sein, da es nach wie vor Teilbereiche und Elemente gibt, die sich nicht allein durch Daten abbilden lassen. Darüber hinaus können Wahrnehmungsverzerrungen (Bias) entstehen, die aus unbewusst falscher Dateninterpretation, unzureichender Datenkontextualisierung oder ungenauer Datenerfassung resultieren. Somit kann den Daten (noch) nicht uneingeschränkt vertraut werden und eine qualitative Überprüfung und Ergänzung ist notwendig.

Im Bereich der Datenanalyse wird die Verwendung von maschinellem Lernen und künstlicher Intelligenz in Zukunft die wichtigste Entwicklung darstellen. Die Fähigkeit von Computern, große Datenmengen schnell zu verarbeiten, technische, taktische und physische Muster auf Einzel- und Mannschaftsebene zu identifizieren und daraus konkrete Handlungsempfehlungen für die praktische Umsetzung in allen Anwendungsgebieten zu liefern, wird dabei von zentraler Bedeutung sein. Fortschritte in der visuellen Objekterkennungstechnologie werden voraussichtlich zur weitgehend automatisierten Datenerfassung führen, sobald eine vergleichbare Genauigkeit und Zuverlässigkeit wie bei der manuellen Erfassung erreicht ist. Die Weiterentwicklung von Teilbereichen wie dem 3D-Modeling ermöglicht es, immer präzisere Daten zu generieren, die Grenzen des heute Vorstellbaren überschreiten und Elemente identifizieren, die vom menschlichen Auge in der Kürze der Zeit kaum wahrnehmbar sind. Fortschritte in der

Datenübertragungstechnologie werden die Latenz zwischen Ereignis und Datenverfügbarkeit verringern und möglicherweise sogar eine Echtzeitverfügbarkeit ermöglichen.

Die potenzielle Anwendung der erwähnten technologischen Fortschritte im Live-Spiel hängt von verschiedenen Faktoren ab. Einerseits spielt die Qualität der Dateninterpretation durch den Dataanalysten eine entscheidende Rolle, andererseits die Kommunikation mit dem Trainerteam auf der Bank. Entscheidender Faktor wird die Bereitschaft zur Umsetzung KI-basierter Vorschläge zur Veränderung der taktischen Ausrichtung, Aufstellung oder Spielprinzipien im In-Game-Coaching durch das Trainerteam sein. Im Rahmen dieses Prozesses muss jedes einzelne Trainerteam durch detaillierte Post-Match-Analysen retrospektiv Prognosen und Handlungsempfehlungen mit dem tatsächlichen Spielverlauf vergleichen, um eigene Erfahrungen zu sammeln und schrittweise Vertrauen in die Technologie aufzubauen.

Infolge des technologischen Fortschritts wird eine Veränderung des Aufgabenprofils des Data-Analysten erwartet. Der Data-Analyst muss zunehmend ein Experte in der Implementierung von komplexen KI- und maschinellen Lernsystemen sein, um (Echtzeit-) Daten für die Analyse von Spielern, Spielen und Gegnern zu nutzen. Die Rolle des Data-Analysten wird sich von der bloßen Bereitstellung und Visualisierung von Analysen hin zu einer aktiveren, aber immer noch überwiegend beratenden Rolle in der Entscheidungsfindung entwickeln. Das bedeutet, dass Data-Analysten bei der Entwicklung und Anpassung von Strategien und Matchplänen im Kontext der Spielanalyse oder der konkreten Spielerauswahl im Scouting stärker involviert sein werden. Es wird zunehmend wichtig sein, dass diese in der Lage sind, Daten, Analysen und daraus gewonnene Erkenntnisse schnell und in einfach verständlichen Formaten zu präsentieren, wie z.B. Visualisierungen, Dashboards und einfachen Worten. Dadurch werden Data-Analysten zunehmend interdisziplinär oder enger mit Trainern, Sportwissenschaftlern und Scouts zusammenarbeiten müssen. Dies wird dazu beitragen, dass diese besser verstehen, wie Daten und Analysen im Kontext der eigenen Spielidee noch effektiver und effizienter genutzt werden können, um die Leistung und den daraus entstehenden Erfolg der Mannschaft oder des Vereins kurz-, mittel- und langfristig positiv zu beeinflussen. Somit werden Datenanalysten in der Lage sein, noch präzisere Prognosen zu entwickeln, die zudem eine höhere Prognosewahrscheinlichkeit aufweisen. Durch die systematische Kombination von Ereignis- und Positionsdaten wird zwangsläufig ein höherer Spielkontextbezug erreicht, wodurch die Validität von Advanced Metrics weiter steigen wird. Daher wird die Mysteriösität der Daten immer weiter abnehmen und die Akzeptanz von deren Nutzen in der Praxis wird flächendeckend zunehmen, weil die Mehrwerte zunehmend unübersehbar werden.

## 8 Literaturverzeichnis

**Backhaus, J.** (2022). Talentförderung, Scouting, Leihgeschäfte. Erfolgsfaktoren für regionale und internationale Vereinskoooperationen im Profifußball. Baden-Baden: Tectum Verlag

**Backhaus, J.; Gaupp, M.; Thaler, E.** (2021). Blocken - Aber richtig!. In: Fußballtraining Dezember 2021. S. 30-39. Münster: Phillipka Sportverlag

**Barthel, M.** (2017). Erkennung und Verfolgung von relevanten Objekten zur semantischen Annotierung von dynamischen, monokularen Szenen am Beispiel von Fußballübertragungen. Doctoral dissertation, Technische Universität München. Online: <https://mediatum.ub.tum.de/doc/1362560/885649.pdf> [abgefragt am: 10.03.2023]

**Bastida-Castillo, A.; de la Cruz Sánchez, E.; Gómez-Carmona, C.D.; Pino-Ortega, J** (2019): Comparing accuracy between global positioning systems and ultra-wideband-based position tracking systems used for tactical analyses in soccer. In: European Journal of Sport Science. Online: <https://doi.org/10.1080/17461391.2019.1584248> [abgefragt am: 10.03.2023]

**bundesliga.com** (2023). DFL: Definitionskatalog Offizielle Spieldaten. Online: [https://www.bundesliga.com/de/bundesliga/news/dfl-definitionskatalog-offizielle-spieldaten\\_0000265687.jsp](https://www.bundesliga.com/de/bundesliga/news/dfl-definitionskatalog-offizielle-spieldaten_0000265687.jsp) [abgefragt am: 10.03.2023]

**bundesliga.com** (2023a). Clubübersicht Saison 2022-2023. Online: <https://www.bundesliga.com/de/bundesliga/clubs> [abgefragt am: 15.02.2023]

**Chen, H.; Schumaker, R.P.; Solieman, O.K.**; (2010). Sports Data Mining. First Online Springer Publishing Company

**DFL** (2020). Wie werden die Offiziellen Spieldaten erhoben?. Online: <https://www.dfl.de/de/innovation/wie-werden-die-offiziellen-spieldaten-erhoben/> [abgefragt am: 28.02.2023]

**DFL** (2023). The source of all – the Official Match Data. Online: <https://www.dfl.de/en/topics/match-data/official-match-data> [abgefragt am: 10.03.2023]

**Fbref.com** (2023). Online: <https://fbref.com/en/> [abgefragt am: 15.02.2023]

**Gruber, F.; Zeilinger, L.** (2021). Fußball das komplette Training für Torhüter. Aachen: Meyer & Meyer Verlag

**Hohmann, A.; Lames, M.** (2005). Sportspiel aus trainingswissenschaftlicher Sicht. In:  
**Hohmann, A.; Kolb, M.; Roth, K.** (Hrsg.). Handbuch Sportspiel (S.132-146). Schorndorf:  
Hofmann Verlag

**Knutson, T.** (2018). Introducing Goalkeeper Radars. Online:  
<https://statsbomb.com/articles/soccer/introducing-goalkeeper-radars/> [abgefragt am:  
11.03.2023]

**Krebs, S.** (2008). Anforderungsprofil des Torwarts im Hochleistungsfußball – Eine Analyse der  
Europameisterschaft 2008 in Österreich/Schweiz und der Bundesligasaison 2007/2008.  
Diplomarbeit. Deutsche Sporthochschule Köln. Online: [https://docplayer.org/32000160-Steffen-  
krebs-diplomarbeit-von.html](https://docplayer.org/32000160-Steffen-krebs-diplomarbeit-von.html) [abgefragt am: 11.03.2023]

**Lames, M.** (1991). Leistungsdiagnostik durch Computersimulation. Ein Beitrag zur Theorie der  
Sportspiele am Beispiel Tennis. Frankfurt a. M.: Deutsch

**Lames, M.** (1994). Systematische Spielbeobachtung. (Schriftenreihe „Trainerbibliothek“ des  
DSB-Bundesausschusses Leistungssport, Bd. 31). Münster: Philippka Verlag

**Lawrence, T.** (2018). Introducing xGChain and xGBuildup. Online:  
<https://statsbomb.com/articles/soccer/introducing-xgchain-and-xgbuildup/> [abgefragt am:  
11.03.2023]

**Loy, R.** (1991). Was fordert das Spiel vom Torwart?. In: Fußballtraining April 1991. S. 3-11.  
Münster: Philippka Sportverlag

**Maiwald, C.** (2014). Data Mining im Fußball. Masterarbeit. Universität Koblenz-Landau. Online:  
[https://kola.opus.hbz-nrw.de/frontdoor/deliver/index/docId/799/file/Masterarbeit\\_DataMining.pdf](https://kola.opus.hbz-nrw.de/frontdoor/deliver/index/docId/799/file/Masterarbeit_DataMining.pdf)  
[abgefragt am: 11.03.2023]

**Memmert, D.; Raabe, D.** (2017). Revolution im Profifußball. Mit Big Data zur Spielanalyse 4.0.  
Berlin: Springer Verlag

**Memmert, D.; Rechner, M.** (2010). Das technisch-taktische Anforderungsprofil des modernen  
Fußballtorwarts. In: Leistungssport Ausgabe 40 (2010), S. 32-37. Münster: Philippka  
Sportverlag. Online: [https://www.researchgate.net/profile/Daniel-  
Memmert/publication/233843278\\_Das\\_technisch-  
taktische\\_Anforderungsprofil\\_des\\_modernen\\_Fussballtorwarts/links/09e4150c0a2e22ee0f0000  
00/Das-technisch-taktische-Anforderungsprofil-des-modernen-Fussballtorwarts.pdf](https://www.researchgate.net/profile/Daniel-Memmert/publication/233843278_Das_technisch-taktische_Anforderungsprofil_des_modernen_Fussballtorwarts/links/09e4150c0a2e22ee0f000000/Das-technisch-taktische-Anforderungsprofil-des-modernen-Fussballtorwarts.pdf) [abgefragt  
am: 11.03.2023]

**Perl, J.** (2017). Interview mit Deutschlands erstem Sportinformatiker: Professor, Dr. Jürgen Perl (Fortsetzung). In: Memmert, D.; Raabe, D. (2017). Revolution im Profifußball. Mit Big Data zur Spielanalyse 4.0. S. 68-70. Berlin: Springer Verlag

**Rahimian, P.; Toka, L.** (2022). Optical tracking in team sports. A survey on player and ball tracking methods in soccer and other team sports. In: Journal of Quantitative Analysis in Sports Online: <https://doi.org/10.1515/jqas-2020-0088> [abgefragt am: 10.03.2023]

**Schmalhofer, A.** (2015). Perturbationsprofile im Nachwuchsleistungsfußball. Doctoral dissertation, Technische Universität München. Online: <https://mediatum.ub.tum.de/?id=1273793> [abgefragt am: 03.03.2023]

**Schnabel, G.; Harre, H. D.; Krug, J.** (Hrsg.) (2016). Trainingslehre-Trainingswissenschaft. Leistung-Wettkampf-Training. 3. aktualisierte Auflage. Aachen: Meyer & Meyer Verlag

**Siegle, M.** (2013). Innovationen in der Wettkampfdiagnostik Fußball. Doctoral dissertation, Technische Universität München. Online: <https://mediatum.ub.tum.de/doc/1120488/1120488.pdf> [abgefragt am: 03.03.2023]

**Slater, M.** (2021). From Barnsley to Belgium, the club owners taking on the elite with data, pressing and young players. Online: <https://theathletic.com/2479880/2021/04/03/from-barnsley-to-belgium-the-club-owners-taking-on-the-elite-with-data-pressing-and-young-players> [abgefragt am: 03.03.2023]

**StatsBomb.**(2023). Data Soccer. Online: <https://statsbomb.com/what-we-do/soccer-data/> [abgefragt am: 03.03.2023]

**Stats Perform** (2023).Opta Event Definitions. Online: <https://www.statsperform.com/opta-event-definitions/> [abgefragt am: 03.03.2023]

**Thaler, E.** (2019). Der Torspieler 4.0 ist der elfte Feldspieler. In: Fußballtraining Juni/Juli 2019. S. 42-49. Münster: Phillipka Sportverlag

**transfermarkt.de** (2021). Behnham – rational & visionär. Wissen ist Macht: FC Brentford mit Moneyball beim 3. Versuch in die Premier League. Online: <https://www.transfermarkt.de/wissen-ist-macht-fc-brentford-mit-moneyball-beim-3-versuch-in-die-premier-league/view/news/381060> [abgefragt am: 28.02.2023]

**transfermarkt.de** (2023). Online: <https://www.transfermarkt.de/bundesliga/startseite/wettbewerb/L1> [abgefragt am: 15.02.2023]

**Wunderlich, P.** (2023). Data Analytics und Data Science. Online:

<https://taod.de/resources/blog/data-analytics-und-data-science> [abgefragt am: 28.02.2023]

**Wyscout.** (2023). Wyscout Glossary. Online: <https://dataglossary.wyscout.com/> [abgefragt am:

28.02.2023]

**Ziegler, M.** (2021). EURO 2020: Torwart-Analyse. Online: [https://www.dfb-akademie.de/euro-](https://www.dfb-akademie.de/euro-2020-torwart-analyse/-/id-11009625)

2020-torwart-analyse/-/id-11009625 [abgefragt am: 11.03.2023]

## 9 Abbildungsverzeichnis

Abbildung 1. Von der Spielanalyse1.0 zur Spielanalyse 4.0. Quelle: Memmert & Raabe, 2017, S.8 .....	8
Abbildung 2: Wechselseitiger Zusammenhang der Gegenstandsbereiche der Trainingswissenschaft. Quelle: Schnabel et. al., 2016, S.19 .....	13
Abbildung 3: Systemtheoretische Gliederung des Fußballspiels. Quelle: Siegele, 2013, S. 23,24,26.....	14
Abbildung 4: Anwendungsgebiete von Data Analytics im Prozess der Generierung von Trainingszielen aus Wettkampfverhalten. Quelle: eigene Darstellung (modifiziert nach Lames, 1994, S.30).....	16
Abbildung 5: Datenarten im Fußball Quelle: eigene Darstellung .....	19
Abbildung 6: Torspielerspezifische Daten und Advanced Metrics Quelle: eigene Darstellung...28	
Abbildung 7: Player Cluster Quelle: eigene Darstellung .....	38
Abbildung 8: Regressionsanalyse grafische Darstellung Quelle: eigene Darstellung .....	39
Abbildung 9: Torspielertypen nach % von KPI Quelle: Eigene Darstellung .....	40
Abbildung 10: KPI Zielverteidigung Quelle: eigene Darstellung.....	41
Abbildung 11: SoTA/90 vs. Save% Quelle: eigene Darstellung.....	42
Abbildung 12: PSxG/SoTAp.90 vs. PSxG+/- p.90 Quelle: eigene Darstellung.....	43
Abbildung 13: PSxG/SoTA vs. PSxG+/- Quelle: eigene Darstellung .....	44
Abbildung 14: PSxG/90 vs. PSxG+/-/90 Quelle: eigene Darstellung .....	44
Abbildung 15: Save% vs. PsxG+/-/90 Quelle: eigene Darstellung.....	45
Abbildung 16: adjustierte Fangquote vs. PSxG +/-/90 Quelle: eigene Darstellung .....	45
Abbildung 17: adjustierte Fangquote Quelle: eigene Darstellung .....	46
Abbildung 18: Adjustierte Fangquote II mit Berücksichtigung von SoTA/90 Quelle: eigene Darstellung .....	47
Abbildung 19: Clean Sheets/90 vs. SoTA/90 Quelle: eigene Darstellung .....	48
Abbildung 20: Clean Sheets/90 vs SoTA/90* Schussqualität Quelle: eigene Darstellung.....	49
Abbildung 21: KPI Mitspielverhalten bei gegnerischem Ballbesitz Quelle: eigene Darstellung ..	49

Abbildung 22: proaktive Verhinderung von gegnerischen Großchancen vs. adjustierte Fangquote Quelle: eigene Darstellung.....50

Abbildung 23: durchschnittliche Distanz der defensiven Aktionen vs. #OPA/90 Quelle: eigene Darstellung .....51

Abbildung 24: Speed vs. Sprints + intensive Läufe/90 Quelle: eigene Darstellung .....52

Abbildung 25: Errors leading to opponents shots Quelle: eigene Darstellung .....53

Abbildung 26: Flanken/90 vs. Stop % Quelle: eigene Darstellung .....54

Abbildung 27: KPI Mitspielverhalten im eigenen Ballbesitz Quelle: eigene Darstellung.....55

Abbildung 28: erfolgreiche Pässe/90 vs. Passquote Quelle: eigene Darstellung .....56

Abbildung 29: durchschnittliche Passlänge vs. Passversuche/90 vs. Passquote Quelle: eigene Darstellung .....57

Abbildung 30: Passquote nach Passlänge Quelle: eigene Darstellung .....57

Abbildung 31: erfolgreiche Pässe nach Passlänge Quelle: eigene Darstellung .....58

Abbildung 32: Passversuche nach Passlänge Quelle: eigene Darstellung .....58

Abbildung 33: Einbindung der Torspieler ins Angriffsspiel Quelle: eigene Darstellung .....59

Abbildung 34: Live Touches vs. Receiving/90 Quelle: eigene Darstellung .....60

Abbildung 35: Ballkontakte nach Spielfeldmitteln/90 Quelle: eigene Darstellung .....61

Abbildung 36: Progressive Distanz bei der Ballkontrolle Quelle: eigene Darstellung .....62

Abbildung 37: Progressive Passdistanz vs. totale Passdistanz/90 Quelle: eigene Darstellung..62