IS-501

Master Thesis in Information Systems

«The genesis of data-driven decision-making in the world of soccer tactics: deciphering the potential of big data»

by Martin Holthe Rønningen

Supervisors

Ilias Pappas & Paolo Spagnoletti

University of Agder, 2021

Faculty of Social Sciences

Department of Information systems

Acknowledgment

This thesis displays research conducted during the spring of 2021 as a final delivery for a Master of Information Systems at the University of Agder in Kristiansand, Norway. The purpose of the master thesis is to assess the utility of a web application constructed to simplify tactical decision-making in soccer. Accomplishing the thesis has been highly instructive and demanding as I had no prior knowledge of programming an application. As a result, the time-consuming and risky endeavor of learning this skill set has resulted in a feeling of mastery.

Furthermore, a special thanks to my supervisors, Ilias Pappas and Paolo Spagnoletti, for their constructive feedback, guidelines, and reflections. I would also like to thank my friend Mads Lochert whose support and knowledge has helped open the door into the soccer analytics community. At last, I would like to thank the enthusiastic respondents for providing rich insight and enlightening conversations by willingly partake in the expert reviews.

It has been two educational and rewarding years at the University of Agder. Thank you!

Kristiansand, Norway

June 4, 2021

Martin Holthe Rønningen

Abstract

Arguably, soccer is a more fluid sport than other popular sports such as baseball, American football, and cricket, which explicitly require more discrete 'plays.' Moreover, with twenty-two players always committed, it has more moving elements than basketball or ice hockey. From a mathematical perspective, this signifies that soccer has more degrees of freedom than these other sports, making it difficult to evaluate the game using one or a small number of metrics. Over time, this challenge has been made exceedingly more difficult within the soccer community. The availability of data has increased much more rapidly than the scientific advancements required to valorize these data. In the aftermath, most recent research papers have elaborated on specific in-game data separately and how to approach them - not how soccer clubs can assemble and utilize them through big data analytics. In accordance, an exploratory case study conducted in collaboration with a major European soccer organization exhibited their state-of-the-art technology to mainly focus on player profiling supplementary to player recruitment - leaving the business in need of a more tactical advancement. To these ends, this dissertation approaches the socio-technical difficulties on how these player-centric metrics can be utilized to simplify tactical decision-making in soccer.

Pursuing the research objective, the author follows a design science research (DSR) paradigm to develop a web application for tactical decision-making. To assemble and test the web application, the author completed an extensive study consisting of two phases. *In the first phase*, a conceptual framework, based on the literature, served as a set of design principles to define both analytical and technological requirements for constructing the web application. Then, in order to fuel the system with valid information, the author studied proven metrics through a systematic literature review and empirical data collected from the major European soccer organization. *In the second phase*, a data-driven system based on polar charts is introduced as a provider of adjustable pre-constructed templates, processing raw data from StatsBomb, to create personalized data visualizations identifying given prospects and tactical patterns according to preferences. Then, a selection of omnifarious soccer experts performed an expert review similar to a black-box test (use case) replicating Tottenham Hotspurs FC and Burnley FC's last match (02/28/2021) to demonstrate the web application's utility.

Drawn from the consensus among experts, the author concludes the system to have shown great potential in generalizing the strategic process of identifying tactical patterns. Additionally, these results strengthen the practical significance of how efficient the artifact is to locate a proper strategy. For example, it took the respondents - unaware of their actual reality - approximately 20-30 minutes to assess and assemble a game-plan almost congruent to 'the special' Jose Mourinho. Furthermore, as the author believes this domain an uncharted territory, the study paves the way towards a digital transformation of sustainable big data solutions for soccer tactics that potentially can generate business value in the future.

Despite a promising outcome, as with any contribution of this type, the sole intention of the proposed web application is to serve as a blueprint for future work. As a result, researchers can practice their discretion to vary what is proposed or submit and achieve improvements. In accordance, an implication extending to the necessity of contextualizing performance indicators on the end user's premises arises as some metrics tend to contradict personal opinions. Additionally, it would be interesting to look further into how the artifact could adapt to reality and which features to be proven in future research.

Keywords: Soccer, Data-Driven Decision Making, Design Science Research (DSR), Big Data Analytics, Web Application, Expert Review, Digital Transformation, Business Value

TABLE OF CONTENTS

1. Introduction **1**
1.1 Background - how one man's bad math helped ruin decades of English soccer 2
1.2 Problem definition - the challenging aftermath of Charles Reep 2
1.3 Research objective 3
1.4 Outline of the thesis 3
2. Theoretical foundation **4**
2.1 Theoretical background 4
2.2 Related works 6
2.2.1 Event - and tracking data (types of data) 6
2.2.2 Coach- and scouting data (utilization of data) 8
3. Research method **10**
3.1 Research philosophy 10
3.2 Research design 11
3.3 Research approach 13
3.3.1 Design science research as a strategy 13
3.3.2 Methodological approach: expert review 15
3.3.3 Qualitative process 17
3.4 Data collection and analysis 18
3.4.1 Phase 1: pre-study 18
3.4.2 Phase 2: main study 18
3.4.3 Method for analyzing data 20
3.4.4 Ensuring the research reliability and validation 21
3.4.5 Limitations 22
3.4.6 Ethics 23
3.4.7 Summary of research methods 23
4. Artifact description: ex-ante evaluation **25**
4.1 Conceptual framework (Design Theory) 25
4.2 Data sources 27
4.2.1 Provider - StatsBomb (FBref) 28
4.2.2 Proven metrics (KPI) 28
4.3 Design and development - constructing the analytical stack 29
4.3.1 Ingestion layer – extract data via pipeline into storage solution (ETL) 31
4.3.2 Data modeling layer – organizing data on top of the big data storage solution 31
4.3.3 Processing layer – number-crunching the proven metrics 33
4.3.4 Data Visualization layer – deciphering the potential of big data 33
4.4 Visualization method (Polar Chart) 34
4.5 Demonstration of use - user scenario 35

5. Findings: ex-post evaluation **40**
5.1 Overview of central themes and consequential errors 40
5.2 Data-driven culture 41
5.2.1 Two-way communication 42
5.3 The respondent's current approach to soccer tactics 43
5.3.1 Quantity of accessible tools 43
5.3.2 Utilization of analytical advancements 43
5.3.3 Suitability for tactical decision-making 45
5.4 Utilization of the artifact 46
5.4.1 Applying a conventional strategy to the artifact (strategy) 46
5.4.2 Reasoning behind the strategic decisions 47
5.5 Practical significance 49
5.5.1 Current approach vs. artifact 49
5.5.2 Comparison and validation of line-ups 50
6. Discussion **51**
6.1 Theoretical implications 51
6.2 Practical implications 54
7. Conclusion and Future Work **56**
8. References **57**
Appendices **61**
A – Expert Review 61
B – Expert Review Software 63
C – Artifact Description 64

LIST OF FIGURES

Figure 1: The systematic literature review's search process. 4

Figure 2: The thesis research design based on an idea by Dube and Robey (1999). 13

Figure 3: DSR Framework based on Hevner, March, Park, and Ram (2004, p. 80). 14

Figure 4: Design science research process based on Peffers et al. (2007). 14

Figure 5: Modes of DSR inquiry retrieved from Sonnenberg and Vom Brocke (2012). 16

Figure 6: Overview of the respondent portfolio, date of the review, and duration. 19

Figure 7: Data Analysis based on Creswell (2014, p. 247). 20

Figure 8: Summary of research method inspired by Dubé & Robey (1999). 24

Figure 9: Rein and Memmert's (2016) conceptual framework. 26

Figure 10: Proven metrics retrieved from FBref.com (StatsBomb). 29

Figure 11: The architectural stack of the web application. 30

Figure 12: Big Data Technologies Life Cycle retrieved from Khan et al. (2014). 30

Figure 13: Florence coxcomb diagram on causes of mortality in the British army. 34

Figure 14: Interpreting the Polar Chart. 37

Figure 15: The prototype. 38

Figure 16: The proposed web application. 39

Figure 17: Aaron Wan-Bissaka - performance overview 52

Figure 18: Heatmaps retrieved from whoscored.com (2021). 54

Figure 19: Attack sides. 54

Figure 20: Shot directions. 54

Figure 21: Average positions. 54

Figure 22: Match report retrieved from understat.com (2021). 55

LIST OF TABLES

Table 1: Concept-centric matrix. 5

Table 2: Types of data sources based on Memmert and Rein's research (2018). 6

Table 3: Expected variables retrieved from Kharrat et al. 's (2020) research. 7

Table 4: Alternative stances on reality (Walsham, 1995). 10

Table 5: Research philosophies 11

Table 6: DSR evaluation retived from Sonnenberg & Vom Brocke (2012). 15

Table 7: Validation criteria based on Guba and Lincoln (1989); Munkvold (1999). 21

Table 8: Overview of the data analysis main findings. 41

Table 9: Respondent's background. 41

Table 10: Summary of technological tools. 43

Table 11: The respondent's tactical groundings. 47

Table 12: Final line-ups compared to reality. 50

1. Introduction

Seldom can the genesis of a novel activity be pinpointed with any accuracy. However, at 15.50 on 18 March 1950, a spectator at Swindon Town's home game against Bristol Rovers took a pencil and notebook out of his pocket. Wing Commander Charles Reep was at that moment starting the first comprehensive notational analysis system for soccer (Pollard, 2002, p. 853).

Although Charles was light-years ahead of his time tracking stats and gathering large data sets, the rapidly growing frontier of new scientific advancements exceeded his notational system. The analogy of prior methods has proceeded into the age of data-driven analytics systems powered by digital information or what scholars refer to as *big data* (Heggernes, 2017; Sykes, 2016). Additionally, Provost and Fawcett (2013) highlight the *data science* domain as the tissue connecting big data processing technologies to a data-driven analytics system. Moreover, the authors define data science as a set of central principles supporting and guiding the extraction of information and knowledge from data (Provost & Fawcett, 2013). Hence, uncovering these large data sets - containing a variety of data types - hidden patterns is called *big data analytics*. In short, the combined use of these terms involves the utilization of facts, metrics, and data to guide strategic decisions that align with every goal, objective, and initiative within an organization. (Heggernes, 2017, p. 136; Provost & Fawcett, 2013, p. 3). Furthermore, as there exist significant variations used to denote big data analytics, recent studies have begun to empirically prove how our choices, behaviors, and even existence in the digital world produce data that offers tremendous opportunities to improve organizational outcomes, current business methods, and practices (Mikalef, Pappas, Krogstie, & Pavlou, 2020; Pappas, Mikalef, Giannakos, Krogstie, & Lekakos, 2018). In comparison, McAfee and Brynjolfsson (2012) presented the potential impact of big data among North American companies already back in 2012 - depending on the data-driven decisions generated from extensive data methods and techniques, companies gained a competitive advantage and outperformed the traditional approaches by 5%. From here on, it became clear that the analytical aspect of big data practices could provide organizations with a competitive edge by creating an information-based arsenal. In sum, Mikalef et al. (2020) suggest that in order to derive value from big data, firms must identify areas within their business that can benefit from the data-driven insight, strategically plan, execute, and bundle the resource mix necessary to turn data into actionable insight. As a result of embracing this kind of high-level analytics to their big data ecosystems, businesses can uncover hidden information, helping their organizations to thrive on data-driven decisions (Pappas et al., 2018; Russom, 2011).

More recently, these expanding opportunities to collect and leverage digital data have functioned as a catalyst for the increasing use of *data-driven decision-making* tools in soccer - a chamber of software and service tools for converting data into actionable intelligence and insight. In brief, all these processes, approaches, and technologies are often referred to by the umbrella term *Business intelligence* (BI) (Arnott, Lizama, & Song, 2017). Moreover, the entrance of such sophisticated tools has led managers over the last decades to alter their decision-making - relying less on intuition and more on data (McElheran & Brynjolfsson, 2017).

In soccer jargon, all the abovementioned data are termed *metrics. B*ased on a vast pool of aggregated data, players are benchmarked against their competitors, where a final *metric* indicates their level in a given player-centric key performance attribute (KPI) (Memmert & Rein, 2018). This process of generating a graphical representation of player performance allows teams to identify an individual's probability of recreating historical actions across thousands of in-game events (Memmert & Rein, 2018). For example, one of the most common metrics assessing an offensive player's match contribution is expected goals (xG) versus actual goals (expected goals difference). A contextual indicator that either reflects if a player underperforms in front of the goalie or if he/she is a great finisher continuously producing a positive correlation in expected goals difference (Kharrat, McHale, & Peña, 2020).

1.1 Background - how one man's bad math helped ruin decades of English soccer

Continuing on Charles Reep's story, Sykes (2016) conceptualizes Reep's analytical approach to "*how one man's bad math helped ruin decades of English soccer."* By segmenting each event data without reflecting on the underlying factors leading up to each outcome, Charles had managed to create a cautionary tale of the damage done when stats go wrong, as England flamed out at the Euro in 1992, playing a misguided style (Sykes, 2016).

After Charles's died in the early 21st century, a far more successful but similar method forced its way into another sport. Led by Billy Beane, the Oakland Athletics went on a 20 winning streak in Major League Baseball (Lewis, 2004). A crucial part of their success rested on how the organization applied a selection system powered by algorithms and statistical data - known as *sabermetrics* or, in vernacular, *Moneyball*. The method identified and filtered suitable prospects based on the Oakland Athletics' specific style of play. Moreover, Neyer (2017) defined sabermetrics as *"the statistical analysis of baseball data ... aiming to quantify baseball players' performances based on objective statistical measurements"* (Neyer, 2017, p. 1). As a result, the successful implementation of the 'sabermetrics approach' is perceived as a turning point in the sport management industry, catalyzing business intelligence applications into the sector (Rangaiah, 2020).

1.2 Problem definition - the challenging aftermath of Charles Reep

In the aftermath of Charles's telling and the realization of sabermetrics, a diversity of new technologies have been added to the rapidly advancing frontier of data-driven practices in soccer (Goes et al., 2020). Each approach strives to make it possible for non-expert users to exploit cutting-edge analysis to their data (Memmert & Rein, 2018). Correspondingly, researchers have been exploring data points separately in this vast pool of opportunities, applying all kinds of techniques to decipher their potential (Anderson & Sally, 2013). In retrospect, an ongoing and timely challenge as researchers of other domains *"found systematic evidence that putting data into action through analysis ... significantly higher productivity in a wide range of manufacturing settings"* (Brynjolfsson & McElheran, 2019, p. 28).

As we know it, the beautiful game ripe for an alteration from conventional, rather qualitative analysis methods to modern data-driven game analysis techniques (Memmert & Rein, 2018). We are starting to recognize how data-driven decisions are changing soccer dynamics, from pre-shaping tactics to pinpointing transfer targets (Anderson & Sally, 2013). As a result of dealing with this trend, Goes et al. (2020) claim modern match analysts require knowledge across the *computer science* and *sports science* domains. Whereas, to make innovative contributions in the future, it seems reasonable that tactical analyses will be increasingly performed by big data technologies across these domains (Memmert & Rein, 2018). For now, this thesis explorative pre-case study supported Goes et al. 's (2020) view reviling most data-driven approaches to be player-centric, as the management is reluctant to implement data completely into their team's game strategy. Arguably, a finding that implicitly emphasizes the lack of proven tactical software in prior research.

Nevertheless, tactical identifiers (metrics) exist, such as press height, passes per defensive action, action zones, usage rates for individual players, or pitch areas (StatsBomb, 2020). However, as management is uncertain of these stats, most tactical work is traditionally done by the coaching staff, limiting both their team analysts and external researchers from exploring new tactical disruptions (Akkermans & van Helden, 2002). Further, Memmert and Rein (2018) justify big data analytics most significant advantage not solely to lie in the magnitude of the underlying data, but the potential depth of insight it provides when assembled across various sources. Hence, the most conventional challenge in data science for soccer appears to be how teams can assemble and utilize proven player-centric tools (or metrics) in collaboration to support the team's tactical decision-making processes.

1.3 Research objective

Throughout the initial sections, the author has elaborated on professional soccer's entry into the computerized world. As mentioned, a timely challenge, as data availability within the soccer community has increased much more rapidly than the tactical advancements required to valorize these data (Goes et al., 2020). A statement leading the author to believe this is uncharted territory, and a study is needed to obtain a deeper understanding of big data's potential impact on soccer tactics. Thus, the overarching aim of this thesis is to contribute to the socio-technical challenges of how data analytics can aid the tactical decision-making processes in soccer by proposing the following research question:

Research question:

How can player-centric metrics be utilized to simplify tactical decision-making in soccer?

Pursuing the research question, the author follows a design science research (DSR) paradigm to develop a web application for tactical decision-making. To assemble and test the web application, the author completed an extensive study consisting of two phases. *In the first phase*, a proposed framework, developed by Rein and Memmert (2016), served as a conceptual framework (design theory) guiding the web application's architectural stack (Figure 11). Next, the author studied theory (proven metrics and conventional systems) through a systematic literature review and empirical data collected from two semi-structured in-depth interviews with a major European soccer club that adopted big data.

In the second phase, the study proposes an artifact for tactical analysis in soccer, in which an expert review demonstrates the web application's proof-of-use.

1.4 Outline of the thesis

The outline of this thesis is as follows: section 2 discusses related works on tactical data-driven decision-making in soccer and emphasizes the overarching gap in current knowledge. Then, section 3 explains the research methodology. Moving forward, section 4 elaborates on the web application's construct and utility, whereas section 5 presents essential findings from experts reviewing the application. Further, section 6 discusses the strengths and limitations of the work in relation to the theoretical background in section 2 and the results presented in section 5. At last, a conclusion is drawn, and further research is suggested in section 7.

2. Theoretical foundation

This thesis's theoretical foundation extends research conducted from a preliminary literature review and an empirical explorative pre-case study. These initial studies were conducted in parallel and investigated the use of big data analytics in a major European soccer organization. In accordance, both studies examined the following research questions:

"How can soccer analytics be utilized in decision-making processes to enhance soccer tactics?"

And

"How has prior literature addressed this?"

Pursuing these questions has been fundamental to comprehend which factors lead the frontier of big data analytics (BDA) initiatives in soccer. Hence, the theoretical foundation of this thesis represents a descriptive theory that informs the initial prototyping of the artifact portrayed in section 4. Therefore, the first sub-section presents an overview of the theoretical background in a concept-centric matrix, whereas the second sub-section elaborates on these leading initiatives according to tactical decision-making in soccer.

2.1 Theoretical background

The reviewed articles listed in the concept-centric matrix below (see Table 1) were compiled based on the knowledge that informs both the research questions in sections 1.3 and 2 - as the author perceives them as complementary. Hence, a systematic review of relevant literature was conducted according to the structured approach provided by Kitchenham et al. (2009); Webster and Watson (2002). Furthermore, the literature search is organized to meet the thesis selection criteria in order to extract the most relevant articles conformed to big data and data-driven decision-making in soccer tactics. As a result, this led to significant findings of proven metrics (Figure 10) drawn from the different concepts illustrated in Rein and Memmert's (2016) conceptual framework (Figure 9). In addition, the inclusion criteria excluded studies not being peer-reviewed, written in English, and published before 2010 in a well-known IS conference, IS journal, book, or other favorable journals validated in the NSD register. To select 20 appropriate publications, the author read through each edition's scope, excluding those who lack relevance, before saving the most fitting candidates in a 'summary of articles' for further assessment. Additional studies to consider were then identified by forward and backward searching the reference lists of included papers. The literature search process is illustrated in figure 1:

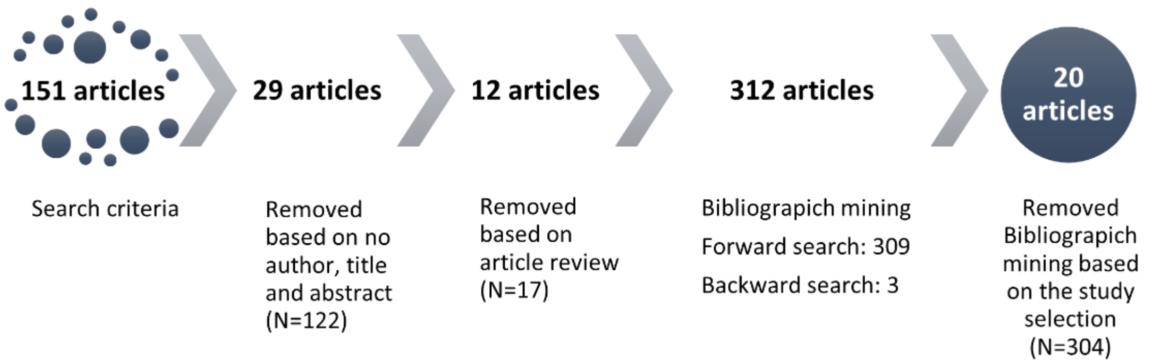


Figure 1: The systematic literature review's search process.

Ultimately, Rein and Memmert's (2016) conceptual framework served as an outline when organizing the matrix (Webster & Watson, 2002). In accordance, each concept (data sources) inside the matrix reflects the most recent developments in data recording technologies that can perform science-based analyses of tactical relevance. The most evident metrics drawn from within these data sources, how to harvest them, and how each layer of the proposed stack contributes to the final visualization are discussed throughout section 4.

Table 1: Concept-centric matrix.

Concepts / **References**	**Data sources**														**Big Data Solution**			
	Coaching & Scouting data							Event data & Tracking data							Architectural layers			
	Offensive Attributes	Defensive Attributes	Roles/Key Players	Player Recruitment	Heat Maps & Pass Maps	Chemistry	Tactical Behavior	Chance Creation	Physics	Critical Success Factors	Player Ratings	On-ball/Technical	Off-ball/Spatial	Expected Values	Architecture	Machine Learning System	Tactics-model	Visualization & Reporting
Khan et al.															✓	✓		✓
Rein & Memmert	✓	✓		✓			✓	✓	✓							✓	✓	✓
Rein & Memmert	✓	✓			✓		✓	✓		✓						✓	✓	✓
F. Goes et al.	✓	✓					✓									✓	✓	✓
Pappalardo et al.											✓					✓	✓	✓
McHale & Relton			✓				✓	✓		✓	✓					✓	✓	✓
Behravan et al.			✓		✓		✓	✓			✓					✓	✓	✓
Hassan et al.							✓			✓	✓					✓	✓	✓
Pappalardo et al.										✓						✓	✓	✓
Rein et al.	✓	✓			✓		✓	✓	✓							✓	✓	✓
Peralta et al.	✓	✓		✓			✓	✓				✓	✓			✓	✓	✓
Erraissi & Belangour															✓			✓
Kharrat et al.			✓	✓						✓	✓	✓	✓	✓		✓	✓	✓
Oussous et al.															✓	✓		✓
Spearman	✓		✓	✓			✓						✓	✓		✓	✓	✓
Van Haaren	✓	✓	✓	✓	✓	✓	✓									✓	✓	✓
Decroos et al.	✓		✓	✓							✓	✓		✓		✓	✓	✓
Dick & Brefeld	✓	✓			✓		✓	✓			✓					✓	✓	✓
Nelli															✓	✓		✓
Llana et al.	✓	✓					✓						✓	✓		✓	✓	✓

2.2 Related works

The related works aim to map out the most applicable metrics for profiling a player according to tactical decisions. Although the overarching concepts in table 1 are divided into three sections (coaching & scouting data, event data & tracking data, and architectural layers), it is pivotal to note how they all intervene to some degree. For example, Spearman (2018) writes about tracking data called off-ball values. Such off-ball values include both offensive and defensive attributes, making them relevant as coaching- and scouting data considering how to utilize them, and tracking data as they give context according to how they are algorithmically processed. Furthermore, to address how they vary from each other, they follow the matrix structure.

2.2.1 Event - and tracking data (types of data)

According to Memmert and Rein (2018), there exist two primary data sources that enable an analyst to gather, capture and contextualize all the various events addressed throughout this chapter:

Table 2: Types of data sources based on Memmert and Rein's research (2018).

Event stream data	Optical tracking data
Event stream data annotates the times and locations of specific events (e.g., passes, shots, and cards) that occur in a game.	Optical tracking data records the player's locations and the ball at a high frequency using optical tracking systems during games.

To obtain digitalized statistics derived from these technical advancements, soccer organizations gather and re-engineer data through self-developed mechanisms or acquire software through well-established stats providers like Opta, Wyscout, Second Spectrum, STATS, SciSports, and StatsBomb. Unfortunately, due to the optical tracking system's expenses, tracking data is only available in wealthy leagues or clubs (Memmert & Rein, 2018).

In accordance, one of the most significant challenges pointed out by Rein and Memmert (2016) is the accessibility to the data types mentioned above. Moreover, a rising privacy issue as commercial institutions, private clubs, and public research institutions oversee the accumulated data logs. This gap is also mentioned as an extension of a professional soccer team's reluctance to share data concerning a possible forfeit in competitive advantages. In addition to filling this gap, Pappalardo et al. (2019) unleashed one of the most extensive open collections of soccer logs ever released. The log contains spatio-temporal events (shots, passes, fouls, etc.) during each match for an entire season of seven dominant soccer leagues (Premier League, La Liga, etc.). To identify which event data are turning out to be the most pivotal from a similar soccer log to Pappalardo et al., Hassan, Akl, Hassan, and Sunderland (2020) applied a neural network analysis tool. The paper resulted in a disruptive approach predicting the most significant sensitivity attributes affecting a match by objectively quantify the match-attributes sensitivity. The analysis outcome identified what the authors deemed to be the 17 most critical performance attributes - out of 75 - to win a game. The most prominent attributes retrieved to win a game from their analysis are *distance covered in the final third* and *pass attempts in the final third*.

Further, Kharrat et al. (2020) undertake the idea behind the performance measures of *expected variables*, as conventional approaches simplicity lack context and a deeper understanding of the situations in which actions are committed. Therefore, the article presents an enhanced methodology based on the conventional plus-minus differential method to calculate expected events vs. actual events. The new method's outcome correlates to a player's performance due to what is expected of him in every context of his gameplay - to see if he underperforms, overperforms, or neither. The method is also intended to predict future performance, likewise, for team performance. Hence, they present the underlying variables illustrated in table 3:

Table 3: Expected variables retrieved from Kharrat et al. 's (2020) research.

Expected goals (xG)	Expected points (xP)	Plus-minus rating
The probability of every shot being a success or goal.	Based on each match difference in xG. The highest xG equals 3 points, lowest 0 points, and similar 1 point.	A modified plus-minus rating, measuring manpower, home advantage, and recent performance.

In accordance with the expected variables, Kharrat, McHale, and Peña (2020) argue that xG is more informative than actual goals when judging how well a team has played. However, since goals are rare, they do not always reflect a team's performance on the pitch. Hence, the perhaps most intriguing outcome and prominent finding are when Kharrat, McHale, and Peña's (2020) model (plus-minus rating) identifies one of world football's hottest properties. Already the season before Ajax's historic run in the UEFA Champions League, their method featured Matthijs de Ligt as a top-five defender in the 18/19 UCL season. Furthermore, due to their re-engineering work with the expected variables, the researchers have begun to partner with a Premier League club, using the same ratings to identify talent across Europe and aid its player acquisition.

Moving forward, Llana, Madrero, Fernández, and Barcelona (2020) express all the abovementioned statistics as so-called on-ball metrics, widely used and accepted by clubs to enhance decisions. However, on the horizon is a new metric taken under the magnifying glass. Llana et al. (2020) define it as off-ball advantages and describe the metric as situations when a player is in a favorable disposition to receive a potential pass. In the case of receiving it, the player would likely improve the possession's value. However, due to the limited nature of event data, these metrics measure a small portion of what actually occurs during a soccer game. To make this concrete, one can consider that, during an ordinary match, Barcelona (now Atletico Madrid) striker Luis Suarez has the ball for a marginal 90 seconds of the 90 plus minutes of game-time. Therefore, what Suarez, or any other player, contributes to the play, like pressing, runs to open space, or tactical positioning, cannot entirely be measured by event data alone (Peralta Alguacil, Fernandez, Piñones Arce, & David, 2020).

Therefore, Spearman (2018) aims to evaluate the quality of the prevalence off-ball metrics discussed above. For example, one considers a tall and unmarked center-forward positioned at the far post during a corner kick. Sometimes the cross comes in, and the center-forward heads it in effortlessly. Other times, the cross flies over his head. Another example is a winger played onside while making a run past the defensive line. Sometimes the through-ball arrives; other times, the winger must break off their run because a teammate has failed to deliver a timely pass. In both circumstances, the attacking player has created an opportunity even if they never received the ball. Hence, the paper constructs a probabilistic physics-based model that uses spatio-temporal player tracking data to quantify what they define as *off-ball scoring opportunities* (OBSO).

2.2.2 Coach- and scouting data (utilization of data)

To apply such *sensitivity attributes* as extracted by Hassan et al. (2020), the promising *expected values* discussed by Kharrat et al. (2020), and the *off-ball metrics* conducted by Llana et al. (2020); Spearman (2018), McHale and Relton (2018) elaborate that these indicators need to be of use to the managers and coaches. Meaning, the attributes should be utilized to identify appropriate line-ups, key-players and exploit the opposition team's weaknesses. Further, Behravan et al. (2019) elucidate that such complex data (soccer logs) generated by automated technologies should also provide valuable knowledge according to a player's tactical whereabouts. Not to the rigidity of the formation in itself. Therefore, the authors present an automatic particle swarm optimization-clustering algorithm to cluster a data log of player performance centers in different games. As a result, they extracted different soccer roles, revealing behavior within their positional cluster, also known as a player's natural position.

Relative to Behravan et al. (2019) findings, does Pappalardo, Cintia, Ferragina, et al. (2019) introduce the ranking system PlayeRank. This data-driven framework offers an honest multidimensional, and role-aware evaluation of soccer players' performance. Moreover, the authors explain that these data-driven player-performance evaluations are getting more fundamental in the soccer industry daily. However, there is no consolidated and widely acknowledge metric for measuring performance quality in all its facets. There are two reasons for this:

1. **On the one hand**, a wealth of sports companies, and television broadcasters, websites such as Opta, WhoScored.com, Footballslices.com, and Understat.com, as well as the plethora of social media platforms for fantasy soccer and e-sports, extensively use soccer statistics (from various sources) to compare the performance of professional players to increase engagement by critical analyses, scoring patterns, and insights.

2. **On the other hand**, team managers and coaches are interested in analytical tools to reinforce their tactical analysis and monitor player's quality during individual matches or entire seasons. Not least, soccer scouts and performance analysts are continuously looking for data-driven tools to improve the retrieval of talented players with in-demand characteristics. A process based on evaluation criteria that contemplate the complexity and the multi-dimensional nature of soccer performance. In turn, measuring performance alludes to computing a data-driven performance rating that quantifies the quality of a player's performance in each match. Proven a complicated assignment since there is no objective and shared definition of performance quality - the bottom-line being soccer organizations to rely entirely on their own data analyst when rating players.

Thus, Dick and Brefeld (2019) approach these player ratings from a conceptual point of view. They proposed an in-depth RL approach to learn valuations of multiplayer positionings using positional data. They claim their work composes the first purely data-driven approach to read and interpret games and, thus, closes the gap toward computational tactics. For instance, correlations between a dangerousness metric and the traditional performance (event data) indicators like the ones elaborated in section 2.2.1 - playing speed, passing, or expansion of teams - could be used to group historic episodes against an opponent. These groups could then automatically devise strategic insights for this rival. For example, the historic groups could indicate counterattacks as more propitious than slow playmaking or that crosses led to more perilous situations. These insights could then be integrated into the individual game plan and support a manager in his decisions.

At last, another tactical perspective undertaken by Van Haaren (2020) is looking at player chemistry from an observational and a predictive perspective. In turn, this could potentially thrive in combination with Dick and Brefeld's (2019) method, recognizing the best line-up for a particular opposition. The purpose of the observational setting is to observe the chemistry between players who have played together. This setting is relevant for a manager who needs to decide on the best possible line-up for an upcoming match. Further, the predictive setting predicts the chemistry between players who have never played together before, which is a particularly relevant context for scouts assessing the fit of a future signing. However, the authors aim to demonstrate how chemistry metrics can enhance team building, identify appropriate transfer targets, decide on the best possible line-up, etc. As a result of their research, the authors present a method capturing the mutual attacking and defensive chemistry for given duos of players. The approach claims to quantify a player's collective impact on scoring goals and prevent their rivals from scoring goals. They also demonstrated how the chemistry metrics could identify an appropriate transfer target for a club.

Furthermore, Van Haaren (2020) uses a Team Builder that assembles a maximum-chemistry team from a given set of players. For example, their model suggested Hakim Ziyech`s – being an Ajax player at the time – to achieve the highest offensive chemistry with the players from Bayern Munich and Chelsea FC. A year later, the player joined Chelsea and started with some very promising displays before getting two fundamental injuries sidelining him for most of the season.

3. Research method

In essence, this section illustrates the selection of the methods adopted to solve the research problem. According to the Frascati Manual, *"Research is the creative and systematic work undertaken in order to increase the stock of knowledge* (OECD, 2015)*."* Moreover, Hellevik (2002) elaborates that any *"means that serves this purpose belongs to the arsenal of methods."* In turn, the author presents a clear rationale for the selected research philosophy, research design, research approach, and research strategy. Furthermore, the author exhibits the data collection and analysis methods to extract constructive findings established throughout the dissertation's two phases (Figure 2). The ***first phase*** contains methods used in the pre-studies leading to this thesis problem definition and research objective. On that account, a literature review was in parallel with an exploratory case study conducted. In addition, these preliminary studies served as a foundation for identifying proven metrics and an illustrative description of the application's development process. Followingly, the author presents ***phase two***. This phase comprises the data generation methods applied for collecting and analyzing data retrieved from experts testing the system. In deep detail and concentrating on the analyzing stage, the criterion for validity and reliability are also manifested here. In accordance, a brief explanation of occurring limitations is also proposed. Eventually, the author elaborates on his perspective and goals regarding research ethics before summarizing the thesis research methods in figure 8.

3.1 Research philosophy

Researchers may have different research perspectives based on different philosophical paradigms when considering how the research should be conducted and how knowledge should be acquired and developed. Hence, a paradigm is a set of shared assumptions that pertain to which philosophical worldview one has and how one perceives different aspects of reality (ontology). Accordingly, Walsham (1995) describes three distinct views on reality: *external realism*, *internal realism*, and *subjective realism* illustrated in table 4:

Table 4: Alternative stances on reality (Walsham, 1995).

Ontology	
Term:	**Definition:**
External realism	Reality exists independently of our construction of it.
Internal realism	Reality-for-us is an inter-subjective construction of the shared human cognitive apparatus.
Subjective realism	Each human person constructs his or her own reality.

Considering Walsham's (1995) definitions, the author perceives reality as subjective and influenced by previous experiences and knowledge. Thus, the author realizes that the research conducted and the results found are not objective. This research's stance on reality is therefore not necessarily like other human's perception of reality. Ergo, a paradigm also represents a perception of how knowledge on reality can be acquired (epistemology). Accordingly, Oates (2006) describes *positivism*, *interpretive*, and *pragmatism* as the three most central philosophical paradigms. These elemental epistemologies emulate the work of Creswell (2014), Goldkuhl (2012), and Oates (2006) in table 5:

Table 5: Research philosophies based on Creswell (2014), Goldkuhl (2012), and Oates's (2006).

Epistemology	
Term:	**Definition:**
Positivism	Research is positivistic when it is characterized by measurable variables, hypothesis testing, and generalization from a small to a larger population (Oates, 2005).
Interpretivism	Interpretive research in IS and computing is concerned with understanding an information system's social context. Concerning the social processes by which it is developed and construed by people and through which it influences and is influenced by its social setting (Oates, 2005).
Pragmatism	Pragmatism is concerned with action and change and the interplay between knowledge and action (Göran Goldkuhl, 2012). According to pragmatism, society is in a continuous process of action. Pragmatic researchers want to create constructive knowledge based on action and are also helpful for future actions. Action is essential for the pragmatic paradigm, not only for the sake of the action itself but also because an action is a path to change. The pragmatic paradigm places the research problem at the center and uses all approaches to understand the problem (Creswell, 2014).

Considering Creswell (2014), Goldkuhl (2012), and Oates's (2006) philosophical views, this thesis's overarching purpose is to address the socio-technical challenge of how soccer analytics can be utilized in tactical decision-making processes. Therefore, the main target is to assess/test the utility of the web application presented in figure 16 and compare the tool to conventional methods applied for soccer tactics. For these reasons, the author could structure the research paradigm under an interpretivism worldview. However, as pragmatic researchers are mentioned to create constructive knowledge based on action, and action is the path to change, the author deems this paradigm more appropriate to a study that intervenes in the world. Hence, from a subjective realism viewpoint, the author perceives the pragmatic paradigm well suited for the undertaken design science research as it interferes with organizational changes and artifact development.

3.2 Research design

In accordance with the author's philosophical stance, the research design is a plan for how the research is carried out from start to finish. According to Yin (2017), a research design is a systematic order connecting empirical data to a study's preliminary research questions and conclusions. Moreover, de Vaus (2001) states a research design to ensure that the data retrieved enables us to explain the initial question as explicitly as possible. Therefore, when constructing the study design, the researcher should reflect on one crucial question (de Vaus, 2001):

"What evidence is imperative to convincingly address the research question?"

Research question:

How can player-centric metrics be utilized to simplify tactical decision-making in soccer?

Subsequently to de Vaus's finding, Jacobsen (2015) elaborates that in order to identify crucial evidence that addresses the research question, the research design should be divided into two conditions depending on whether the study is broad (*extensive*) or in-depth (*intensive*), and whether the study is descriptive or explanatory (Jacobsen, 2015). Correspondingly, this study has chosen to go in-depth, using design science research as a research strategy and an expert evaluation as a data generation method.

Furthermore, researchers can distinguish between describing or explaining (Jacobsen, 2005). A *descriptive design* involves describing a situation that leads to a rich and detailed analysis of a specific phenomenon and its context. The analysis tells a story that includes discussing what has happened and how different people perceived what happened. On the contrary, when using an *explanatory design*, the researcher goes further than a descriptive design. The aim is to explain why things happened the way they did and whether notable outcomes occurred. In addition, an attempt to identify several, often coherent factors that have affected an event's outcome is made (Oates, 2006; Jacobsen, 2005).

Based on the above arguments, this study has both an explanatory and a descriptive design. A preliminary literature review and an explanatory pre-case study were first carried out to argue for what has become the application's research context, construct, and contribution. Hence, the research design has also been twofold. In phase 1, the author inherits the perception of a practitioner-oriented framework (Figure 9) for introducing computer-driven systems in soccer and, in parallel, aggregates a systematic literature review, which led to a concept-matrix of critical success factors (proven metrics). This part of the study is partly theory-building as the framework is built and aggregated based on several sources from research and practice. Then the author applied the criteria from the framework to conduct an exploratory case study. The purpose was to identify whether the proven metrics were taken into account in one of Europe's largest football clubs - to reveal challenges in today's systems. Based on findings from *phase 1*, the author developed a web application for tactical analysis to propose concrete measures that can handle the identified challenges elaborated in section 1 and thus increase the prerequisites for a successful game plan. *Phase 2* has an explanatory design. The author wants to explain how soccer clubs can handle the tool by proposing a concrete proof-of-concept through a scenario demonstrating the application's usability and an expert review assessing its utility.

At last, the action plan in figure 2 illustrates the research design that enables the author to collect and process data needed to pursue the research question:

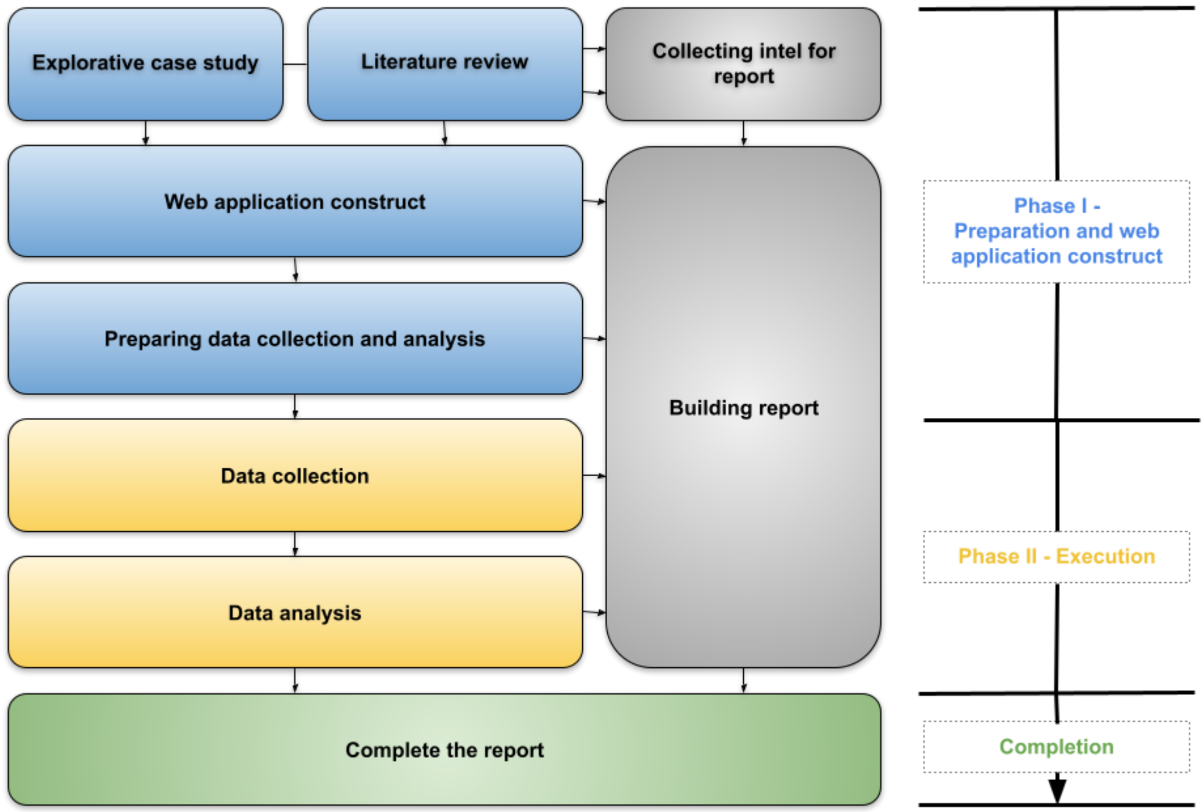


Figure 2: The thesis research design based on an idea by Dube and Robey (1999).

3.3 Research approach

Ultimately, to present conclusions that can address the research problem, the study has comprised design science research as a strategy with a qualitative approach to evaluate data retrieved from the chosen research method of expert review.

3.3.1 Design science research as a strategy

To convincingly address the research question in line with traditional IS research, the author follows Hevner et al.'s (2004) framework (Figure 3) for design science research (DSR). The DSR strategy aims to solve practical challenges and gather knowledge so that the phenomenon can be further researched (Hevner, 2014). The solutions to these challenges and problem areas are acquired through the development and application of an artifact. Moreover, artifacts can be constructions, models, and prototypes that contribute to IS research so practitioners can more easily understand and solve identified challenges connected with IS implementations (Hevner, March, Park & Ram, 2004). Therefore, the design research process involves several steps, and it often ends in a series of changes in the product (see Figures 14, 15, and 16). In this thesis, the artifact is a web application for tactical analysis in soccer.

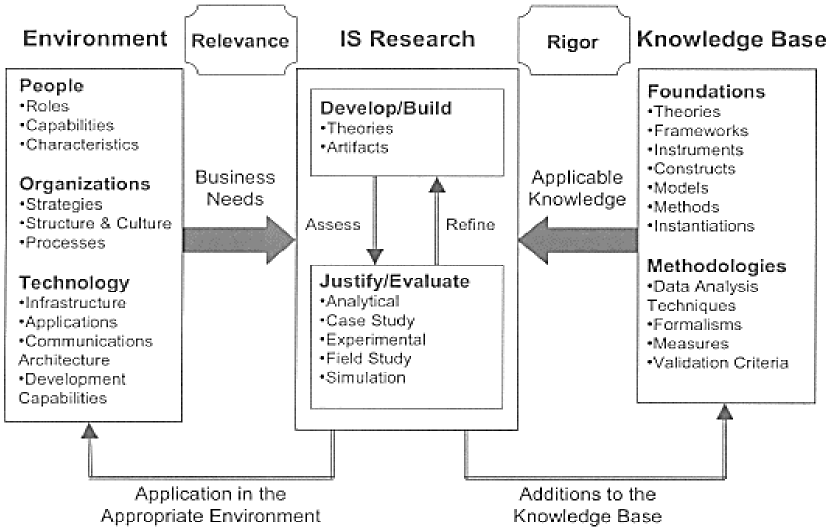


Figure 3: DSR Framework based on Hevner, March, Park, and Ram (2004, p. 80).

In brief, the framework consists of two concepts. First, we have the *relevance concept*, which focuses on what is essential in organizational practice and what feeds this thesis's particular business need (see Section 1). Second, to assess this particular need, one applies the *rigor concept*, which focuses on what is known about the problem and what relevant knowledge to apply (see Section 2). In turn, combining these concepts into the evaluation, development, and *building cycle* of the web application, Hevner et al. (2004) claim the approach to help solve the business need – which in this case, is the lack of holistic tactical soccer tools. Correspondingly, the undertaken research utilizes an expert review to assess, *justify*, evaluate, and refine the proposed web application's utility throughout the cycle.

In essence, Gregor and Hevner (2013) explains the DSR method to match that defined by Peffers, Tuunanen, Rothenberger, and Chatterjee (2007), and the following steps are incorporated into this thesis structure: (1) *identifies problem*; (2) *define solution objectives*; (3) *design and development*; (4) *demonstration*; (5) *expert evaluation*; and (6) *communication*. At the same time, constituting other approaches, Peffers et al. (2007) developed a synthesized research process model illustrating the steps above to involve five activities, which is adjusted in relevance to this thesis structural context in figure 4:

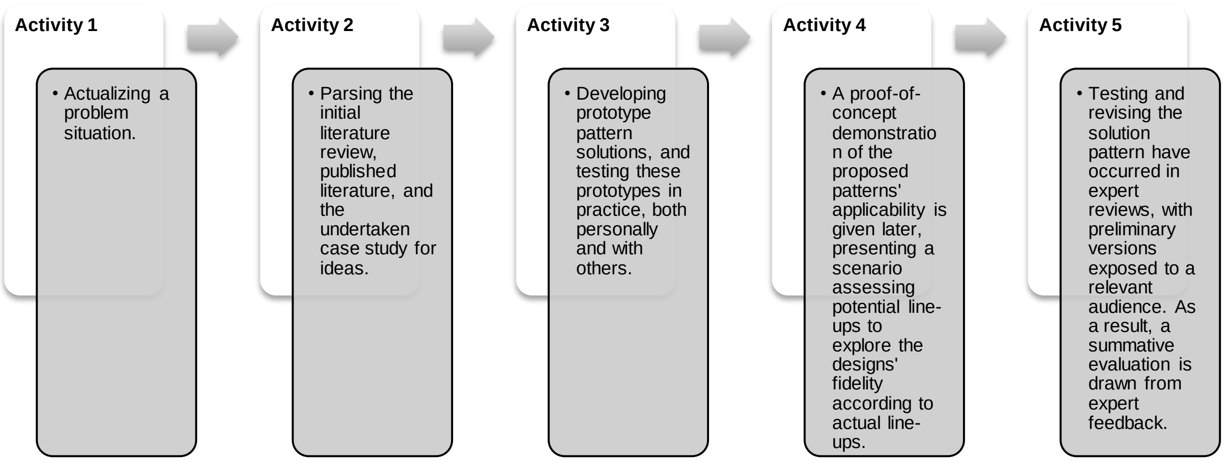


Figure 4: Design science research process based on Peffers et al. (2007).

3.3.2 Methodological approach: expert review

In order to prove the utility of an artifact, Horváth (2007), Baskerville and Myers (2015) explain "*The dual mandate of the DSR: (1) to utilize the gained knowledge to solve problems, create change or improve existing solutions; and (2) to generate new knowledge, insights, and theoretical explanations.*" For that reason, researchers in DSR must apply proper methodological inquiry types within behavioral theories that provide specific direction for the chosen procedures (Creswell, 2014). Building upon this classification, Fischer, Winter, and Wortmann (2010) elaborate that there is a broad consensus that kernel theories are theories from natural or social sciences that serve as a foundation for artifact construction (Walls et al. 1992, p. 42). However, as the author has taken a pragmatic stance, Goldkuhl (2004) and Venable (2006b) consider kernel theories as only one way of grounding and point out the importance of artifact impact over artifact grounding. Hence, the author has chosen the methods of inquiry to reflect more upon the characteristics and research strategy stipulated in section 3.3.1.

Moreover, this allows for adding to the following qualitative research process (see Section 3.3.3) by utilizing an expert review as a methodological approach for validating and exploring the artifact's usefulness. In accordance, the author can ask managers if they find the proposed artifact useful when evaluating the novel web application according to the given organizational context in sections 4 (interior mode) and 5 (exterior mode). To these ends, this thesis mirror Sonnenberg and Vom Brocke's (2012) three principles for evaluating the web application illustrated in table 6:

Table 6: DSR evaluation retived from Sonnenberg & Vom Brocke (2012).

Principle	Description
Distinction between **interior and exterior modes** of DSR inquiry	This principle directs the foci of evaluations on two aspects: (1) the constituents of the artifact and the design decisions taken as well as on (2) the evaluation of the usefulness of the artifact.
Documentation of prescriptive knowledge as **design theories**	This principle necessitates the prescriptive knowledge to be documented in a structured way. This would facilitate the communication and dissemination of the prescriptive knowledge produced within a DSR process. Moreover, such documentation would already have a truth-like value that is worth to be accumulated in a DSR knowledge base.
Continuous assessment of the DSR progress achieved through ex ante and ex post evaluations	This principle prompts the design researcher to have multiple evaluation episodes throughout a *single iteration* of a DSR process.

Furthermore, as Sonnenberg & Vom Brocke (2012) argue that these principles are interrelated in that one principle supports the other principles, the author has applied Sonnenberg & Vom Brocke's (2012) modes of DSR inquiry illustrated in figure 5. This method builds on prior work on DSR evaluations. It extends the notion of *ex-ante* - and *ex-post evaluations* by emphasizing that in order to achieve rigor in DSR, it is not sufficient just to let the IT artifact emerge in the build phase and evaluate its use, but to ensure proper design decisions in order to consistently and rigorously converge to a feasible and valuable artifact. In particular, it is argued that by following these principles and methods, the prescriptive knowledge produced in section 4 (ex-ante evaluation) can be considered to have a truth-like value proving the artifact's utility.

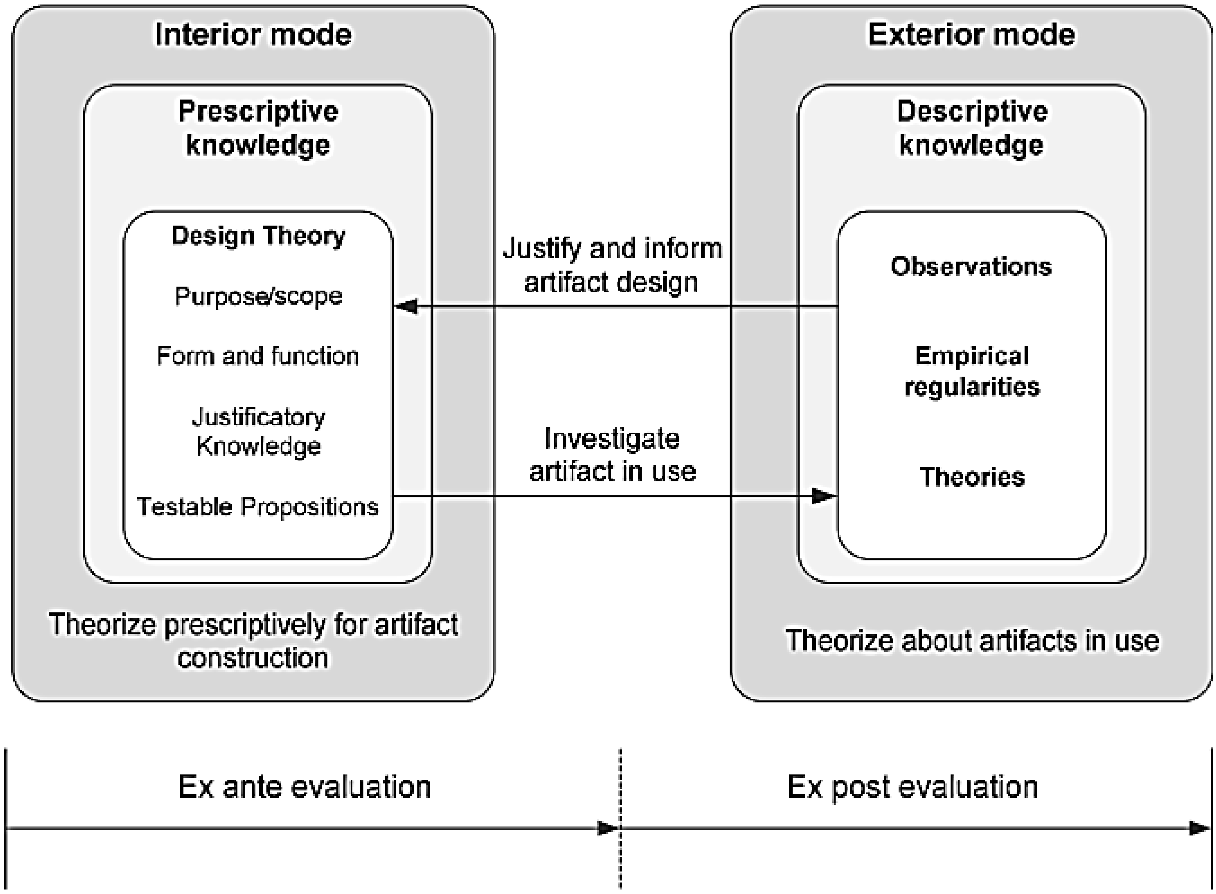


Figure 5: Modes of DSR inquiry retrieved from Sonnenberg and Vom Brocke (2012).

As a result of demonstrating an ex-ante evaluation (design theory) in section 4, the unique phenomena (see Section 1 and 2) that appear from the synergy of people, organizations, and technology would be qualitatively assessed through an expert review in section 5 (ex-post evaluation). According to Sonnenberg and Vom Brocke (2012), maintaining this 'build-evaluate'-like pattern embodied in existing DSR methodologies would have substantial epistemological implications on the validity of knowledge created while the artifact emerges. Hence, the author can generate knowledge adequate for further theory development or problem-solving (Klein & Myers, 1999).

3.3.3 Qualitative process

In line with the author's pragmatic stance, strategy, and methodological choice, a qualitative study provides a researcher with opportunities to examine cultural and social phenomena in IS research (Myers & Avison, 2002). Moreover, Oates (2005) characterizes the qualitative approach as examining few events in-depth to gain broader knowledge and understanding about a phenomenon. Hence, as holistic-tactical-analysis in soccer is an area where there is little previous research, the author considered it appropriate and necessary to acquire this in-depth understanding of how professionals interact with the web application. As a result, the qualitative approach encourages the author to comprehend and rationalize the occurring interplay between knowledge and action drawn from experts testing the artifact. Additionally, data collection techniques such as an interview also aid in structuring and processing qualitative data, documents, and participatory observation (Myers & Avison, 2002).

Furthermore, it is essential to note that the composed artifact is a component of a human-machine problem-solving system. For such developments, knowledge of empirical work and behavioral theories is crucial to construct and assess them. The constructs, components, and methods are therefore exercised within relevant environments by appropriate subjects. Because the proposed artifact represents the 'machine' part of the human-machine system, the principal aim is to determine how well it works, not theorize around or confirm anything concerning why it works (Gregor & Hevner, 2013). Thus, the informants should have a relative background such as coach, data consultants, computer science, or sports science with a footballing background. Determining how well it works would have been more challenging if the study had used a quantitative method - as the method does not consider the context of the study (e.g., scenario) and is less flexible than the qualitative methods. Particularly when the specific social and institutional context forming this research will be lost in the quantified textual data. For example, as the artifact is in a continuous cycle (build-evaluate pattern) of improvements, it is impossible to improve and adjust a survey after being sent out (Neuman, 2014).

To these ends, the data collection needed to be flexible to adjust the interview guide/expert review along the way, as minor defects or lack of contextual data could occur in the system. Subsequently, this proved appropriate as the author - before the study - had little knowledge of how various soccer clubs operate their data-driven systems for tactical use. The author was thus able to continuously adapt and sharpen the study's focus based on new feedback and contextual understanding of how others perceived the application. Based on the characteristics of qualitative approaches and the study's scope, the author believes that the choice of approach has provided a reasonable basis for answering the research question.

3.4 Data collection and analysis

In this section, the author presents the data collection techniques used for this study in-depth. First, in phase 1, the author elaborates on the methods used in the pre-case study leading to this thesis problem definition, the literature review process identifying proven metrics, and the illustrative description of the applications development process in figure 11. Then, in phase 2, the qualitative expert interviews are given as the primary data source. Further, a description of how the author has carried out the data analysis is elaborated, followed by a section assuring the research's quality (reliability and validity). At last, an elaboration of the study's ethical dilemmas and the methodological limitations are presented. Thus, this section explains the key parts of the process that have shaped the author's understanding and provided this study with a qualitative foundation.

3.4.1 Phase 1: pre-study

In the autumn of 2020, a literature review was in parallel with an exploratory case study conducted and used as a foundation for the web application construct - shown in figure 11. Moreover, the conducted systematic literature review focused on scrutinizing existing literary contributions encompassing analytical metrics used in soccer. Based on the most proven metrics, an 'instruction of use' was proposed to guide developers in designing an analytical visualization for strategic decision-making in soccer. Further, in collaboration with two employees of a significant European soccer club, this thesis pre-case study confirmed the 'instruction of use' as feasible in their environment. These participants had different tasks and areas of responsibility. However, both were tightly linked to all strategic decisions based on data.

Further, the pre-study results helped form the basis for this thesis's focus and development of the interview guide in the main study. As a result, the author learned that participants were dissatisfied with the environment's current lack of tactical analysis systems. Hence, the author chose to develop a tool based on these two studies to address this issue and test it during the main study.

3.4.2 Phase 2: main study

In this phase, both interviews and participatory observations were used to collect data and generate findings. Hence, findings are based on an iterative process (hermeneutic process) where developing prototype pattern solutions and testing these prototypes were done personally and with others. Finally, testing and revising the conclusive solution patterns have occurred through expert reviews, with preliminary versions exposed to people familiar with the subject.

Generation method - expert evaluation

Selection of informants

When selecting informants, it is essential to note that the composed application is a component of a human-machine problem-solving system. For such developments, knowledge of empirical work and behavioral theories is crucial to construct and assess them. Therefore, the constructs, components, and methods of the use case (expert review) were exercised within relevant environments by appropriate subjects (Figure 6) relevant to soccer tactics. Additionally, as Goes et al. (2020) claim, modern match analysts require knowledge across the *computer science* and *sports science* domains; one selection criterion was to acquire an in-depth understanding of how professionals from various domains interact with the web application.

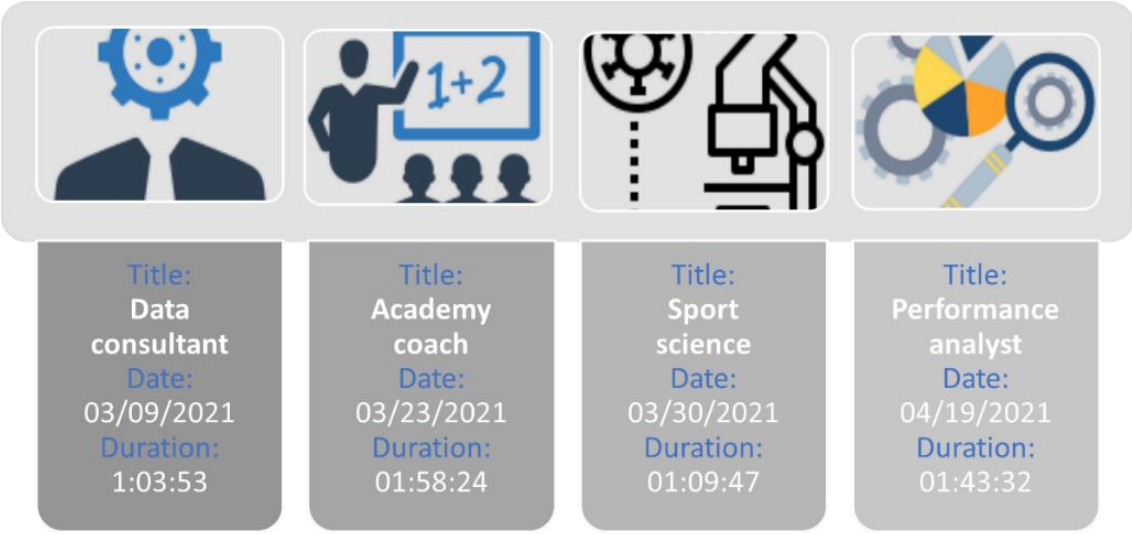


Figure 6: Overview of the respondent portfolio, date of the review, and duration.

Conducting the interview/expert review

The author did not use the same procedure for all the interviewees, as some minor defects were detected and fixed as the informants identified them. However, these minor flaws did not affect the reliability or validity of the review. The interviews were performed with a semi-structured interview guide. To better understand the web application's utility, the interviewees were sent a preparation document (see Appendices A, B, and C) and read the proof-of-concept illustrated through a scenario (see Appendix C). They were also presented with the use case's restrictions to save time on formalities during the test/evaluation. Hence, they had the opportunity to prepare and reflect on the questions in advance. Some of the questions were of the general sort, such as "*How would you specify your current role regarding analytics?*". In contrast, others were adapted to test and evaluate the system, like "*Costumed for your needs, would this artifact simplify your current approach to tactics?"*

In some cases, it was appropriate to let the interviewee steer the direction of the interviews. In that way, the author got into topics that the author would not get into if he had conducted a more structured interview. The author wanted to conduct the interviews as individual face-to-face interviews because it provides solid communication. Most of the interviews were conducted over Zoom or Teams (video conferencing tools) to avoid travel and due to the corona-pandemic. The overview of the informants is shown in figure 6.

In short, the interview guide is divided into four phases. Phase 1 serves as an introduction where the author presented himself and explained the assignment's focus (if needed) and how the interview would take place. Further, phase 2 consisted of some simple background questions and the respondent's current approach to data analytics. Phase 3 represented the utilization of the artifact (expert review) with accompanying questions. Finally, phase 4 was the end of the interview. The author noted a summary of each expert's assessment and opened up for the respondent to correct any misconceptions and provide the necessary additional information.

Observation

The main gathering of data is collected through observations both during the interviews and after. The informants let the author record their interaction with the artifact on video while utilizing and evaluating the application according to the use case.

3.4.3 Method for analyzing data

For the expert reviews, recorded semi-structured interviews were transcribed and analyzed. The data analysis was then performed using thematic analysis by Creswell (2014). The underlying figure represents the hierarchical qualitative data analysis approach. Although the method seems rigid, Creswell (2014) concludes it as more interactive in practice. Furthermore, the approach fits both qualitative and quantitative data analysis when the plan explicitly uses interviews as the source of raw data. The main findings extracted from the data collections expert evaluations are categorized and analyzed according to Creswell (2014) thematic framework and illustrated in table 8, consisting of four expert interviews (Figure 6 and Appendix A):

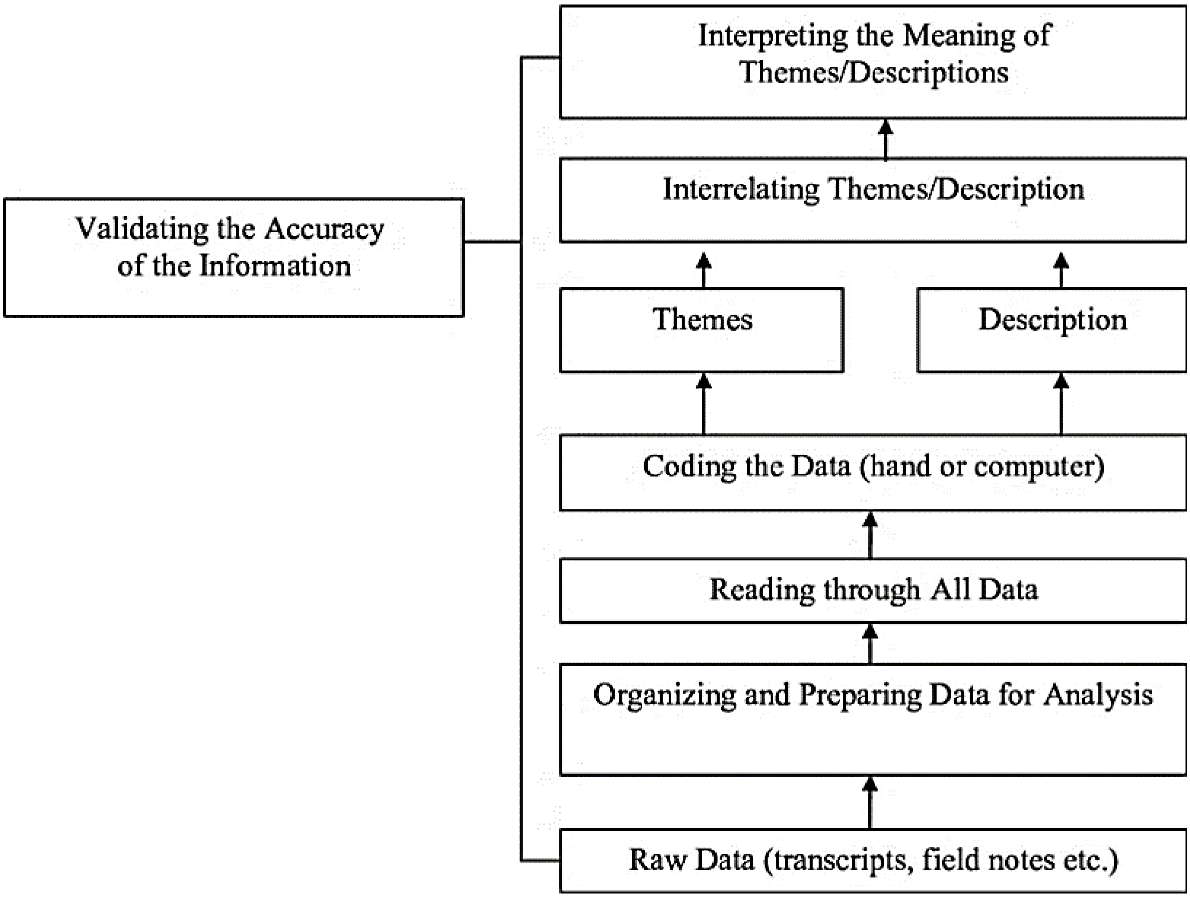


Figure 7: Data Analysis based on Creswell (2014, p. 247).

Thus, when analyzing the information from the expert evaluation, it was essential to transform the raw data into the same format by transcribing the interviews. Moreover, there are two forms of transcription; naturalized and denaturalized (Davidson, 2009; Oliver, Serovich, & Mason, 2005). In naturalized transcription, the interview must be written down words for words and with as many details as possible (e.g., dialect) (Hutchby & Wooffitt, 1998; Oliver et al., 2005). Using a denaturalized approach to transcription, one tries to differentiate the interview's substance, which means the main content, rather than depicting dialect and accent (Davidson, 2009; MacLean, Meyer, & Estable, 2004). The author has chosen to carry out denaturalized transcription in this research, which quickly finds the most pivotal content. Notations were also written down on personal notes based on observations and reflections made under the interviews. Additionally, the author identified the main themes in the data using these segments presented by Oates (2005):

1. Segments that are not relevant to the study.
2. Segments that provide general descriptive information which helps understand the respondent's context and history.
3. Segments that are relevant to illuminate the focus of the study and/or answer the research question.

The categories were then selected based on previous research (deductive) and collected data (inductive). Much of the literature the author studied deals with big data solutions and proven metrics according to various systems and tools. Therefore, it is natural that the categories were chosen partly based on this. Hence, by basing the categories on previous research, such as combining Rein and Memmert's (2016) framework to related works, the author could test whether the results are reflected in research and illuminate areas the research does not cover. For example, the author learned that little research focuses on applying knowledge management in connection with user support. Finally, the categories were placed in table 8 (see Section 5.1). Using a table to structure categories, the author got an overview of connections, shortcomings, and patterns that helped highlight the study's focus. Based on these reasons, the findings were linked to relevant research, and the author could finally draw theoretical and practical conclusions based on the research objective.

3.4.4 Ensuring the research reliability and validation

In order to ensure research reliability and validation, Guba and Lincoln (1989) have developed criteria for assessing the coded data from the data analysis. These criteria are illustrated in table 7 as *credibility*, *transferability*, *reliability*, and *validity* (confirmation):

Table 7: Validation criteria based on Guba and Lincoln (1989); Munkvold (1999).

Criterion	Aim	Strategy
Credibility	Establish consistency between the respondent's constructed reality and perception of reality as presented and attributed to the various stakeholders of the researcher.	Fieldwork and observations over a longer time. Discussion of data and results with external colleagues and informants.
Transferability	Present sufficiently detailed findings to make it possible for the reader to consider if these findings can be transferred into other contexts.	'Thick' elaboration.
Reliability	Ensure that the methodological changes and the interpretive process are documented so that the reader can follow both process and choices made through the path.	Make the research design explicit.
Validity	Ensure that data, interpretations, and results are based on context and not just due to the researcher's imagination.	Make data available. Describe the logic used to transition from data to the results.

Credibility

According to the credibility principle, the author tried to involve and engaged the informants, so the presented data corresponds with the informant's reality. However, this study has been going on for such a brief period. It lacks, therefore, a more extended period of studying big data in a proper soccer organization.

Transferability

Most of the interviews conducted are rich in information and then in details, nuances, and variations. Hence, the author has tried to present a 'thick' description of the findings.

Reliability

One cannot expect another researcher to develop the same results if the same study is done again. Therefore, following the reliability criterion, interpretations and decisions have been documented to be followed, making the research design explicit for the reader.

Validity

Under the criterion of validity or confirmation, the author has made data available by publishing the transcribed interview in appendix A. The logic behind the encoded data is described in section 3.3. The reader can trace back to the source in the form of the interview guide. The results do not just come from personal interpretations; the transcribed interviews are also attached to this report.

3.4.5 Limitations

In this section, the limitations of the author's research approach will be discussed. In general, a standard limitation of qualitative research is the small scale of conducted interviews. Nonetheless, as the characteristics of a qualitative approach are to examine few events in-depth to gain broader knowledge and understanding about a phenomenon, one can argue for a more prosperous and granular data collection. Moreover, as mentioned above (see Section 3.4.2), the author applied a selective sampling technique relying on bias judgment when choosing whom to participate. Hence, in addition to volunteer bias, the sample is also prone to errors of judgment by the researcher, and the findings would not necessarily end up being representative. Ultimately, this could be seen as a possible limitation as more representative respondents might not have been identified.

There are also limitations to managing an interview individually, as a 'research associate' enables a bilateral critique of each other's execution, which might lead to a more in-depth interview (Walsham, 1995). Furthermore, employees in modern organizations regularly are extremely busy and pressured. For example, in one of the interviews, the respondent expressed a shortage of time for the interview, signifying that the data gathering of this event could have been more granular. However, it is considered more beneficial to finish the interviews and suffer some interaction-time if the respondents are pressured, making the probability of re-meeting the organization on another occasion bigger (Walsham, 2006).

At last, in order to address some of these limitations, the author presents data directly from the transcribed interviews, allowing the readers to evaluate the data and determine the adequacy of the author's reasoning.

3.4.6 Ethics

Shared and mediated by benevolence for this study, the author obtained a great deal of information from four semi-structured expert reviews. Thus, as a researcher, the author must act in a 'correct' and ethical way to safeguard the participant's values. Violation of ethical guidelines would not only harm the author's reliability as a researcher but undermine the confidence of the whole research community at the University of Agder. Restoring such trust may be overly time-consuming and limit other researchers from doing their work (Israel & Hay, 2007). That being said, it has been highly prioritized to behave morally trustworthy as a scientist. In addition, the author assured all engaged participants to make an informed consent to cooperate, so they could understand the meaning of the research and then willingly consent to join (Israel & Hay, 2007).

Further, the author acquainted all participants with the research objective and reassured them that the data would be processed anonymously. Hence, when utilizing video conference tools to record the interviews, this was done according to the participant's approval. They were also informed that they were entirely entitled to terminate the interview at any point if desired. Following the interviews, all the participators got the chance to correct statements or highlight expressions they did not want to be quoted.

Additionally, if the interviewees were to give appropriate information, they could not reveal their workplace. Hence, the clubs are anonymous throughout the thesis. Moreover, this is also a reoccurring topic during the interviews. Since data-driven systems debuted, there have not been any actual day-to-day issues concerning ethics or GDPR when utilizing data. For now, it is said that stakeholders do not consider data to be proprietary or anything of the sort. However, in the future, this could very well be the case, as players wanting to own their data in the same way as image rights – currently being paid as parts of contracts in certain leagues.

Finally, with admiration for other researchers and their achievements, the author has done his utmost to avoid plagiarism. In this context, the author has been cautious about crediting other researchers by following the APA 6th standard (Oates, 2005).

3.4.7 Summary of research methods

Figure 8 illustrates a summary of the research design for this study. The left side of the figure shows the overall research approach (design science research) utilized to pursue the overarching research question. The right-hand side shows how the inquiry types are interrelated within each phase of the study. Thus, the methods illustrated in figure 8 parallels that described by Peffers et al. (2008) and include the following color-coded steps: (1) identify the problem; (2) define solution objectives; (3) design and development; (4) demonstration; (5) evaluation; and (6) communication. In combination, empirical data and previous research have formed the basis for discussing findings and drawing conclusions about the study's proposed web application according to the research question in section 1.3.

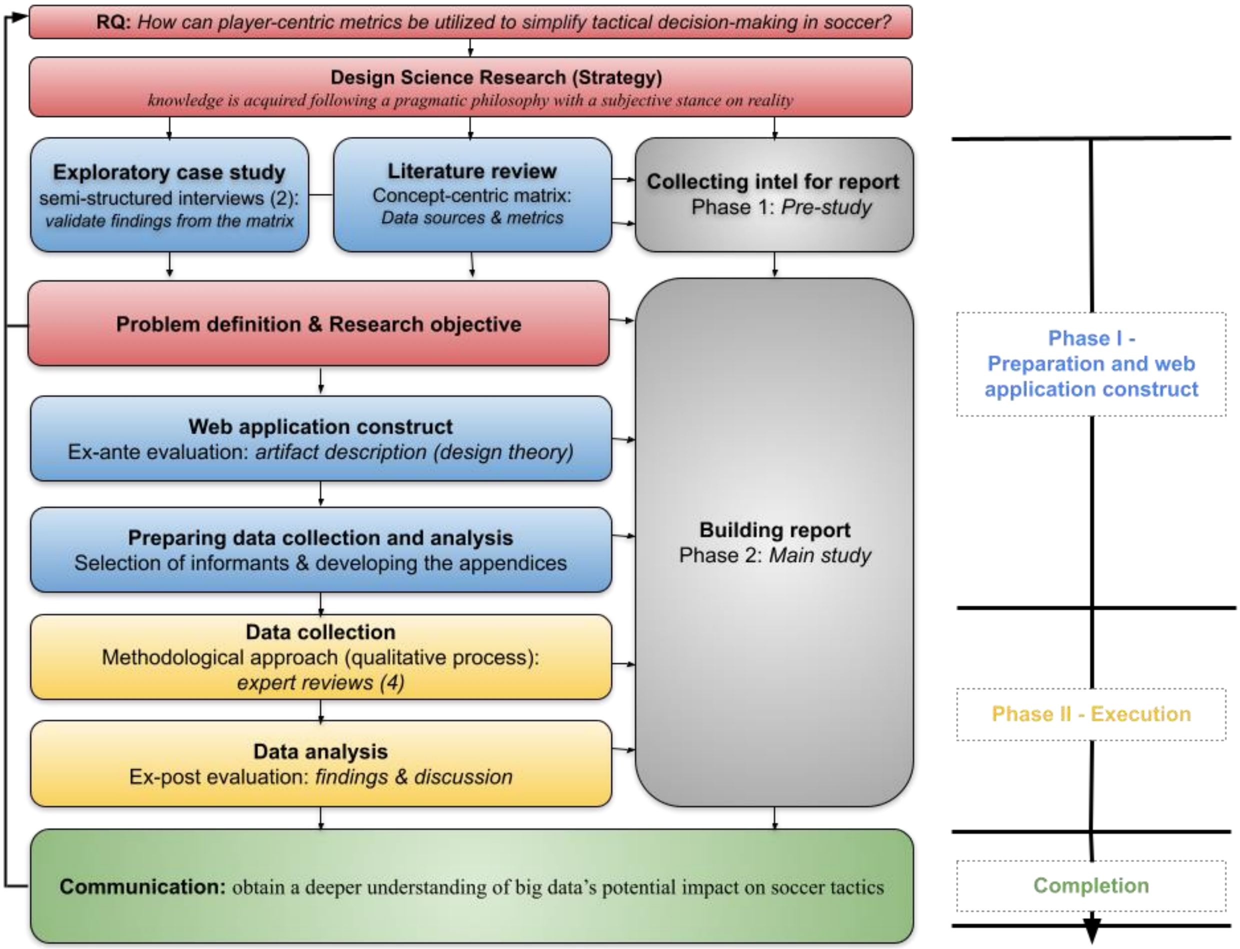


Figure 8: Summary of research method inspired by Dubé & Robey (1999).

4. Artifact description: ex-ante evaluation

To theorize prescriptively for artifact construction, the author has until now established how the underlying potential of data has opened the gateway for soccer analytics. Examining this relatively untouched domain has ultimately formed the design search process (see Section 1-2), guiding this section's prescriptive statements about how the artifact is designed, developed, and brought into being (Sonnenberg & Vom Brocke, 2012). Moreover, for prescriptive knowledge produced in DSR to have a truth-like value, this chapter derives from Sonnenberg and Vom Brocke's (2012) three principles (Table 6) to address the interior evaluation mode of design decisions - reflecting the ex-ante evaluation (interior mode) illustrated in figure 5. To these ends, this chapter documents the design theory (ISDT), demonstrating the artifact's purpose, its rationale, its inner structure, the conditions under which the artifact is expected to work, and the steps required to use the artifact in practice.

Further, as Goran Goldkuhl (2004) and Venable (2006) emphasize the need for ISDTs, but do not insist on a kernel theory-based grounding, the chronological process of forming the artifact extends to Rein and Memmert's (2016) *conceptual framework* (Figure 9). Moreover, the artifact's visual appearance reflects the conceptual framework's final output (visualization and reporting) in figure 9. As the previous literature review and the exploratory pre-case study were conducted through a rigorous and systematic process, the results from these studies were used to inform the web application's main component – The Polar Chart (see Section 4.4 and 4.5). Going forward, the complete version of the artifact merges modified polar charts with a tactical drag-and-drop whiteboard interface (Figure 16). Whereas the polar chart aims to visualize how the inferred statistics or proven metrics described in section 4.2 can mesh into interpretable synergies. In brief, the purpose of the complete version - constructed by utilizing the technological stack presented in figure 11 - is to aggregate an end user's desirable player attributes from the representation in figure 10 to reveal a team's tactical capabilities or playing patterns. The slices within each participating player's chart and the various *data sources* (illustrates a specific attribute) fueling them are described using StatsBomb terminology (The most common standards) in section 4.2.

At last, the theoretical grounding dedicated to construct and showcase a clear contribution of the artifact's utility in a real-world application environment - from which the research problem was drawn - includes the blueprint of the applications *architectural stack, visualization method*, and a *demonstration of use*.

4.1 Conceptual framework (Design Theory)

Considering all the soccer analytics initiatives discussed in the introduction and the theoretical foundation, it can be concluded that none specifically recommends how big data technologies can be used to perform holistic analyses that are science-based and of specific relevance for soccer tactics. However, Rein and Memmert (2016) proclaim to have undertaken the future challenges and opportunities for sports science based on prior research by utilizing big data as a foundation. As a result, they present a framework for researchers to grant complex processing algorithms that allow non-expert users to exploit cutting-edge analysis to their data. A framework that has guided the overarching process of constructing this thesis proposed artifact. Further, Memmert and Rein (2018) assume no other structured approach recommends how big data technologies can perform science-based analyses of practical relevance. In accordance, the authors define big data with regards to soccer as follows:

- **Volume:** describes the data's magnitude and refers to the size of a given dataset in soccer (Rein & Memmert, 2016).
- **Variety:** is distinguished into *structured*, *semi-structured*, and *unstructured data* and is referred to as data heterogeneity - different data formats and data sources consist of a pre-defined schema depicting the data. *Structured data* allows simple maneuvering and searching through the data where a relational database system is a canonical example. In contrast, *unstructured data* deficits a definite schema, with video data and text messages being standard examples. *Semi-structured* data falls amid these two extremes and consists of data that lacks a pre-defined structure but may have a variable schema that is often part of the data itself (Rein & Memmert, 2016).
- **Velocity:** characterizes the data production rate that describes the speed with which novel data is generated. In soccer, the velocity varies widely between real-time streams from physiological and positional data to delayed data from notational analysis during training and competition (Rein & Memmert, 2016).

With big data as the substructure of their framework, Memmert and Rein (2018) elaborate on how an analysis model for soccer should incorporate various *data sources* reflecting the most recent developments in data recording technologies. The model's central purpose is to combine information from the various areas (data sources) to conclude game performance – both individual and team performance.

Further, Rein and Memmert (2016) illustrate the framework as a big data technology stack or a system architecture for soccer tactic analysis organized along with several levels listed on the right side, followed by their guidelines in the following paragraphs:

First, the necessary infrastructure to gather the data is required spanning physiological, psychological, crowd, coaching, scouting, and tracking data in addition to observational data and video.

Second, a storage system is required to allow efficient data storage and access.

Finally, a processing pipeline is established to extract relevant information from the data and subsequently merge it to build an explanatory and predictive model.

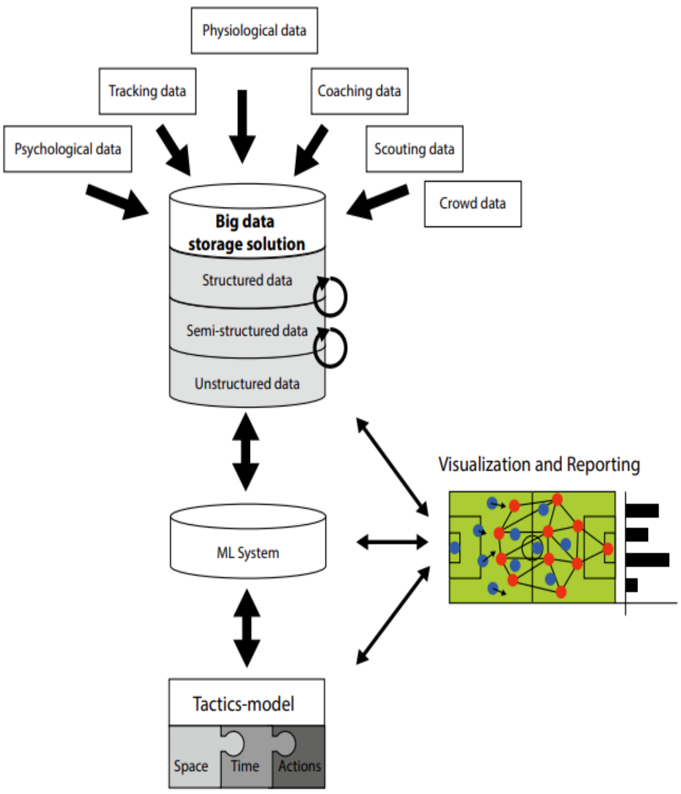


Figure 9: Rein and Memmert's (2016) conceptual framework.

In a modified model of the figure above, Memmert and Rein (2018) explain the data sources to consist of equivalent proportions of raw data and already-processed data that could be stored, for example, in a centrally accessible unit in a configured *data lake* or a *data warehouse* **(**big data storage solutions) - to make queries as simple as possible. Correspondingly, to extend the conventional approach's lack of contextual information, a processing layer must be constructed to extract relevant information from the data. Furthermore, by combining relevant data obtained from processing with the conventional approaches, one can visualize desirable reports by applying suitable algorithms. Finally, a possible solution to somewhat reduce the resulting analysis complexity could be achieved using machine learning methods to generate data-driven models. In short, the system provides contextuality to conventional stats. For instance, shots on target (a conventional stat) do not contribute much to game performance unless they tell the expected result (context) of these shots - like the xG (see Table 3).

4.2 Data sources

As mentioned above, the conceptual framework elaborates the data sources to consist of equal proportions of raw data and already-processed data. As discussed later in section 4.3.2, this information should be stored in a centrally accessible unit in the form of a data warehouse, making queries as simple as possible for the end-user (Memmert & Rein, 2018). For now, there exist two primary data sources that value soccer matches – illustrated in a representation of table 2 beneath:

Event stream data	Optical tracking data
Event stream data annotates the times and locations of specific events (e.g., passes, shots, and cards) that occur in a game.	Optical tracking data record the player locations and the ball at a high frequency using optical tracking systems during games.

To decipher the potential of utilizing these sources, one must first acquire proven data from a legitimate source (McHale & Relton, 2018; Spearman, 2018). For now, there exist companies like *Opta, Wyscout, InStat, Second Spectrum, STATS, SciSports,* and *StatsBomb* generating one or both sorts of these data. Based on the pre-case study, these companies were also viewed as the most legitimate in the industry. Moreover, due to these companies and the resources needed to maintain the most granular data, two significant challenges are pointed out by Memmert and Rein (2016):

- First, the accessibility to soccer data.
- Second, due to the optical tracking systems' expenses, tracking data is only available in wealthy leagues or clubs.

Both challenges are rising privacy issues as commercial institutions, private clubs, and public research institutions oversee the accumulated data logs. This issue is also an extension of the professional soccer club's reluctance to share data - concerning a possible forfeit in competitive advantages. For now, as internal data gathering does not possess the same granularity as leading vendors provide, most clubs attend to retrieve their data by licensing these prominent company's products or APIs (StatsBomb, 2020). An API (application programming interface) is the software intermediary that allows two applications to talk to each other and share manageable information. In sum, the data sources are the applications, databases, and files that an analytics stack integrates to feed the data pipeline described in section 4.3.1.

4.2.1 Provider - StatsBomb (FBref)

When choosing a data source or supplier, there is mentioned in the pre-case study a consensus within the analytics community that StatsBomb features richer data with better quality than competing vendors. Furthermore, it can be seen that StatsBomb data is a direct product of reverse engineering as the vision with StatsBomb data was to transition football data from the world of proxies into a more accurate reflection of what is actually happening on the pitch. For example, in contrast to other suppliers, StatsBomb adds Shot Impact Hight (the degree of difficulty according to the balls Z-coordinate when a shot is taken) to the Expected Goals (xG) algorithms. As a result, this provides a much more realistic picture of the measured event. In sum, StatsBomb data is more comfortable to merge with tracking data than any other event data on the market (StatsBomb, 2020).

Moreover, to decipher the utility range from such informative data, the research compiled across the concept matrix (see Section 2.1) has examined and applied best practices to craft comprehensive metrics for soccer evaluations – so-called proven metrics or key performance indicators (KPI's). Like many of these studies, the metrics used in this thesis are retrieved from FBref, an easy-to-use source for football stats delivering StatsBomb data for the top 5 European leagues (FBref.com, 2021).

4.2.2 Proven metrics (KPI)

Overall, metrics refer to a wide diversity of algorithmic data points indicating performance engendered from a multitude of methods. In general, these metrics are measures of quantitative assessments commonly used for reviewing, benchmarking, and tracking performance (Bose, 2004). Typically, they are grouped and integrated into a holistic dashboard for management or analysts to maintain performance assessments, estimations, and business strategies. Also known as key performance indicators (KPI's) or proven metrics (Peral, Maté, & Marco, 2017). Nevertheless, as every power user has access to these various data sources and metrics, this vast pool of options can make it challenging to choose the appropriate instruments necessary for assessments and evaluations.

Hence, this thesis's most proven metrics have been established by merging the initial systematic literature review's contributions with soccer analysts (interviewed in the pre-case study) real-life experiences. As a result, 24 data points were extracted as the most critical outputs measuring the game's activities. Moreover, to assess and allocate the chosen data points delivered by StatsBomb's most recent technological developments and their respective sources, each concept within the matrix (see Table 1) was used to reflect a phase from which these metrics could originate. In sum, as this resulted in four phases typically known to dictate the game, each metric (KPI) was assigned to its initial phase by color codes - **possession**, **transition**, **attacking,** and **defending** - in the glossary presented below in figure 10:

GLOSSARY

Possession For passing & ball progression	Transition For dribbling & positioning	Attacking For goal-scoring & chance-creation	Defending For defending
Pass Comp % The percentage of attempted passes that successfully reach a teammate	**Turnovers Per 90** Includes every time a player dissposses or miscontrol where the player loses the ball	**Goal Creation-Action (GCA) Per 90** The two offensive actions directly leading to a chance, such as passes, dribbles and drawing fouls	**Rate Adjusted Tackles Won %** The number of successful dribbles a player makes per 90 minute, adjusted for their success rate.
Progressive Passes Per 90 Completed passes that move the ball towards the opponents goal at least 10 yards - excludes passes from the defending 40 % of the pitch	**Succesful Dribbles %** Percentage a successful attempt at taking on a player and pass theme while retaining possesion	**Shot Creation-Action (SCA) Per 90** The two offensive actions directly leading to a chance to shot, such as passes, dribbles and drawing fouls	**Interceptions Per 90** The amount of secsessfully interseption per 90 minutes
Passes into 1/3 Per 90 Completed passes that enter the 1/3 of the pitch closest to the goal - not included set pieces	**Rate Adjusted Dribbles Per 90** The number of successful dribbles a player makes per 90 minutes, adjusted for their success rate	**xG (Expected Goals) Per 90** Expected goals per 90 minute included set pieces such as penalty kicks.	**Ball Recoveries Per 90** Every time a player regain possession of a loose ball per 90 minutes
Passes into Box Per 90 Completed passes that enter the penalty area of the pitch closest to the goal - not included set pieces	**Progressive Distance Per Carry** The number of yards a player carries the ball towards the opposition goal per carry the player complet in the average 90 minutes	**xA (Expected Assists) Per 90** Expected assist Per 90 included set pieces	**Pressure Regain %** Number of times the team win the ball within 5 secunds after the player applyed pressure to opposing player who is receiving, carrying or
Pressured Passes Per 90 Passes made while under pressure from opponent	**Touches in 1/3** The amount of times a player touch the ball in the final third towards the oppenent goal	**Non-Penalty xG + xA Per 90** Non-Penalty expected goal + expexted assist per 90 minute	**Dribbled Past** Number of times a player is dribbled past by an opposing player
	Touches in Box The amount of times a player touch the ball in the opponents penalty area	**Shots on Target %** Percentage of shot attempts that is on target	
	Pass Target Number of time a player was the target of an attempted pass	**Aerials Won %** Percentage of aerial attempts that is succesfully won	

Figure 10: Proven metrics retrieved from FBref.com (StatsBomb) (FBref.com, 2021; StatsBomb, 2020).

4.3 Design and development - constructing the analytical stack

This section reviews the typical structure of an analytical system and how it is applied to the web application. Hence, the process is described in accordance with conventional design theories used to solve similar research problems - as undertaken in this thesis - for other domains than soccer.

Initially, as all analytical artifacts depend on access to data from their relevant sources, there is an urge for technological tools able to fetch and analytically number crunch those data (Gupta & George, 2016). According to Kwon, Lee, and Shin (2014), acquiring a fully functional BDA system fulfilling this pressing urge entails a stack of technologies to address the life cycle illustrated in figure 12. In brief, a stack characterizes a set of rigorous mechanisms and modular technologies, allowing operators to design enterprise-grade and powerful applications (Oussous, Benjelloun, Lahcen, & Belfkih, 2018). Similarly, a data analytics stack incorporates distinct technologies that allow users and firms to develop a robust analytics engine to assemble, merge, clean and transform data from, e.g., StatsBomb (Erraissi & Belangour, 2018). Moreover, the stack consists of several interdependent layers that form an effective and fully operational analytics system, with each layer providing a unique level of processing (Oussous et al., 2018). In addition, and utterly important when addressing a proper solution, is the comprehension of how each layer interacts - as each layer is symbiotic and depends on one another to function within the system. Typically, this co-construction involves a *data warehouse* (storage system) build upon sound *data modeling*. Which, in turn, are contingent on a robust *data pipeline* for ingesting and processing desirable data before it is visualized for strategic decisions (Oussous et al., 2018). The final analytical stack of the web application is illustrated in figure 11.

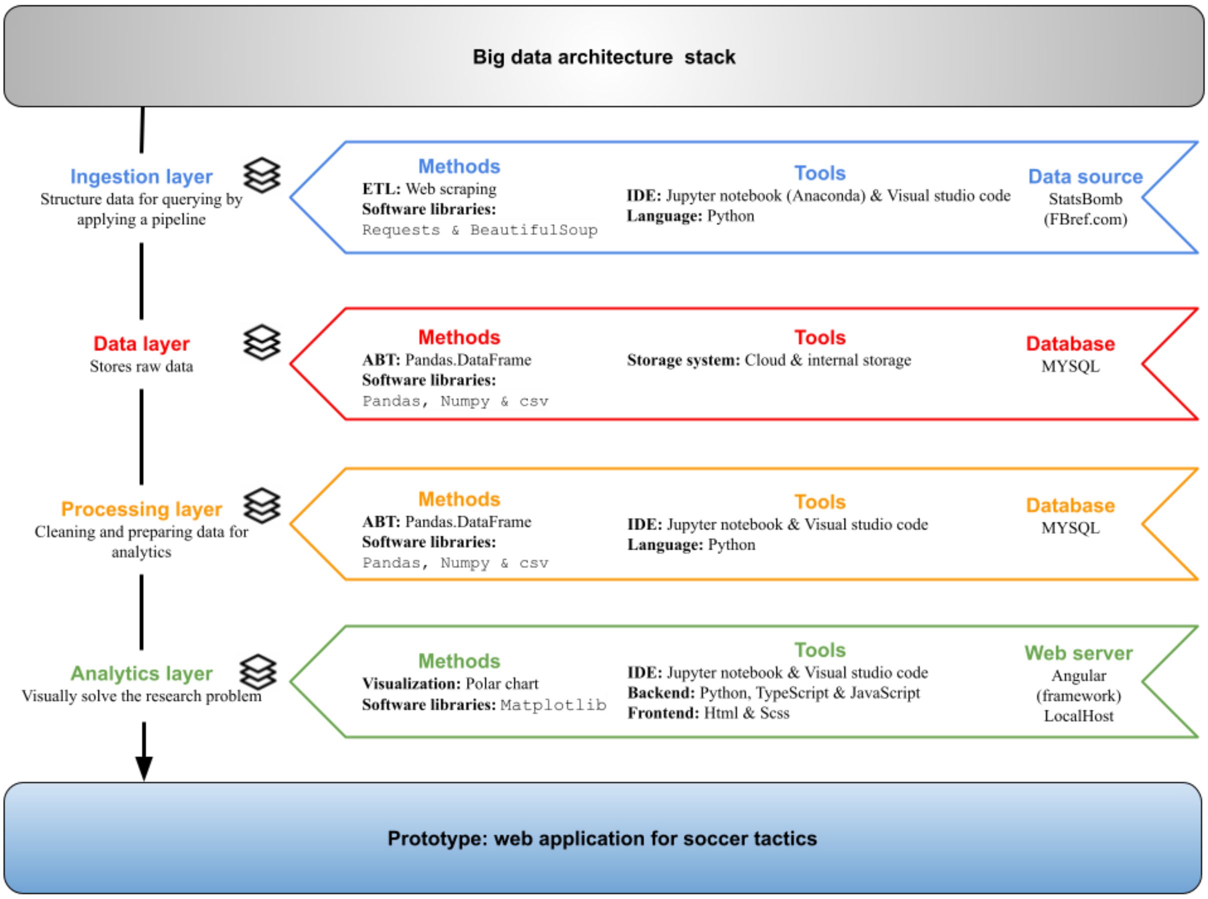


Figure 11: The architectural stack of the web application.

In sum, Ivanov and Singhal (2018) state that distributed big data processing and analytics applications require a complete end-to-end architecture stack comprising several big data technologies. In addition, there are many potential architectural patterns in order to fulfill the web application requirements. Hence, the following sub-sections will elaborate on each layer selected for this thesis BDA architecture, following a similar pattern as shown by Khan et al. (2014) in figure 12 and how it can be applied to soccer.

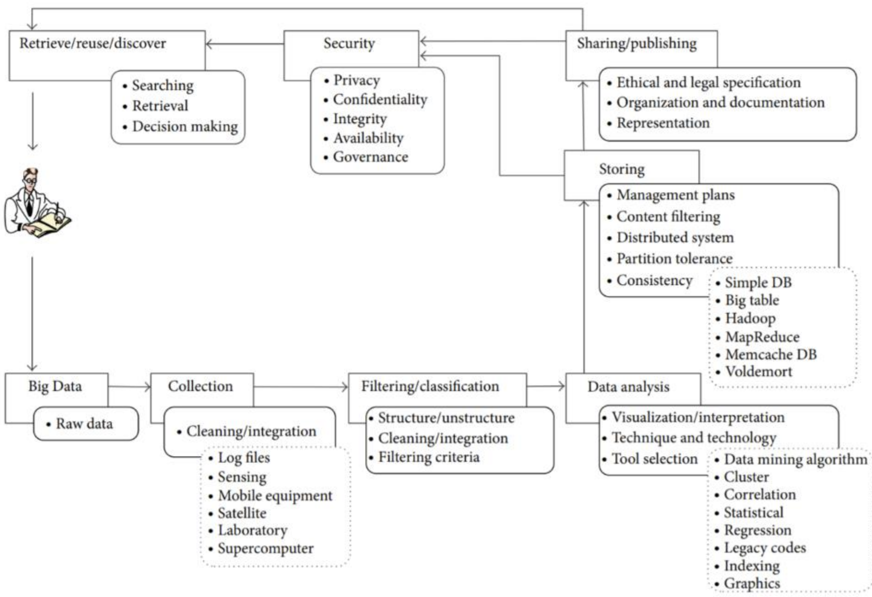


Figure 12: Big Data Technologies Life Cycle retrieved from Khan et al. (2014).

4.3.1 Ingestion layer – extract data via pipeline into storage solution (ETL)

After establishing which data to fuel the system (Figure 10), the significance of the ingestion layer comes into being. The purpose of the ingestion layer is to allocate and integrate chosen data from where it is originated into a layer where it can be stored and analyzed. In a sense, the ingestion layer prepares data for the specialized tools and technologies utilized in the later layers (Erraissi & Belangour, 2018).

Further, to route data into a storage solution, it must be reproduced from an external source. The most common processes for pulling data from its source, alternatively transform it if needed, and push it into a data warehouse are ETL (extract, transform, load) or ELT (extract, load, transform). This process is generally performed as a data pipeline consumes information from external sources and store it in a specific destination (e.g., data warehouses or data lakes). The pipeline represents the software that secures data from a technical perspective and makes it available for strategic use - typically applied to internal analytics and product features. Most pipelines ingest raw data from multiple sources via a push mechanism - e.g., an API call – where a replication engine pulls data at regular intervals. Some of the most common technologies used for this process are Blendo, Stitch, and Kafka, launched by Apache (Palanivel, 2019).

In order to configure a pipeline for data ingestion, there are two general paths: 'do-it-yourself' (DIY) or use a prebuilt tool. If an enterprise chooses the DIY route, data engineers typically use scripting and programming languages such as Python, Ruby, Go, Java, or Bash to construct their custom ETL jobs (Mitchell, 2018). Unfortunately, developing a data pipeline from scratch adds a considerable burden to maintenance and infrastructure development. Nevertheless, as this thesis web application depends on ingested data from StatsBomb, the author has utilized the DIY route to avoid costs associated with licensing an API. In theory, Mitchell (2018) defines *web scraping* as the procedure of mining data through any method other than a program interacting with an API. Hence, as illustrated in the model construct (Figur 11), the author extract, transform and load StatsBomb metrics into tabular data tables stored in a MySQL database via a self-made web scraper (ETL-method) using Python (Mitchell, 2018).

4.3.2 Data modeling layer – organizing data on top of the big data storage solution

After the web scraper has ingested raw data from StatsBomb, the most fundamental layer in the analytical stack is often referred to as the data layer - representing the backend of the entire system. Besides storing all the raw data (Figure 10) from different data sources fed by the pipeline, this layer operates the modeling process that structures and organizes data to support the analytics (Mitchell, 2018). In sum, this process grants users to alter data for selective querying (Palanivel, 2019). For example, a corporation often enables analytical modeling by constructing an analytical base table (ABT). Moreover, this tabular base table's structure is similar to an Excel spreadsheet (similar to CSV files) conceived by aggregated and clean data extracted through the web scraper's pipeline (Nelli, 2015). In accordance, the author utilizes the Pandas library's DataFrame (see Figure 11) in Python to organize the web application's raw data into self-made ABTs (2-dimensional spreadsheets) (Nelli, 2015). In general, this allows data scientists to create, clean, and analyze consistent data, providing better performance and truthfulness.

In addition, after reviewing the exploratory pre-case study (see Section 3.4.1), two conventional approaches for modeling data sources in soccer was clarified:

1. Resourceful clubs apply pre-structured data modeling from external sources to fuel their self-adjusted systems. In some cases, they partially gather data themselves.
2. Minor clubs hire best practice services through data consultancy companies like StatsBomb, Wyscout, Analytics FC, etc.

As abovementioned, the pre-case study informants claim StatsBomb as the most proven and trusted source to retrieve layered soccer data within the analytics community. StatsBomb supports this claim implicitly and advertising that their:

> Unique event data collection spec has over 3,400 events per match of on and off the ball data including pressures, ball carries, possession chains and more. Data generated from a blend of Computer Vision and human-driven collection with automated validation checks and a highly experienced quality assurance team, makes it the most accurate event data in the industry (StatsBomb, 2020).

Big data storage solutions

After managing the data modeling, two challenges arise for a storage solution: (1) providing the imperative infrastructure; and (2) developing appropriate algorithms and processing processes (Rein & Memmert, 2016). Thus, the vital data processing infrastructure must be developed, enabling manageable storage and subsequent access to the gathered data (data entry). Moreover, to store the massive amounts of raw data automatically retrieved from a vendor such as StatsBomb, it is common to use a data warehouse (DW). In short, data gathered through vendor APIs tends to be already-processed data, which is beneficial as a DW is structured for this purpose only. However, there is no such thing as a go-to solution when choosing a big data storage system. One of the pre-case study informants (a data analyst) mentions that as long as clubs are granted access to the correct data and how to filter and interpret data correctly, they have come a long way. Hence, as this thesis web application is a smaller project, not dependent on massive storage, the author stored the ingested data into ABTs in a minor MySQL database ready for later processing. Other standard technologies used in this layer are Amazon S3, Hadoop HDFS, MongoDB, etc. (Oussous et al., 2018).

At last, Decroos, Bransen, Van Haaren, and Davis (2019) elaborate on the importance of structuring a storage unit according to the context it is supposed to enhance, as most existing methods for valuing soccer events suffer from three crucial limitations that future research must face. *First*, these approaches mostly ignore actions other than shots and goals. In accordance, most work to date has concentrated on the concept of the expected value of a goal attempt. *Second*, existing approaches tend to assign a rigid value to each action, regardless of the conditions under which the event is performed. For example, many pass-based metrics treat passes among defenders in the defensive third of the pitch without any pressure. On the downside, they measure passes between attackers in the offensive third under heavy pressure from the rivals equally to unpressured passes. *Third*, most approaches only review immediate effects and fail to account for an action's effects a bit further down the line. In order to address these limitations, the author has focused on structuring the storage unit's various data entries according to the player's positional context on the field (see Section 4.5). Thus, defensive player charts are given easy access to the ideal stats and vice versa for midfielders and attackers.

4.3.3 Processing layer – number-crunching the proven metrics

After the data sources have been allocated by the pipeline and transformed into a desirable ABT stored in the MySQL database, the processing layer starts the actual number crunching (Nelli, 2015). This process is arguably the most crucial in the end-to-end big data technology stack - as analysts process a large volume of data into relevant data marts before the final presentation layer (also known as the business intelligence layer) (Palanivel, 2019). First, a data warehouse or database is used as a centralized unit for holding data from various sources. An enterprise can then transform and model this data and then build visualizations based on analytics and business intelligence software. Similarly, when number-crunching StatsBomb metrics into suitable algorithms for the web applications position-based player charts, the author utilizes the NumPy library in Python for efficient processing (Nelli, 2015).

However, enterprises are now moving to cloud data warehouses to take advantage of their scalability and reduced maintenance overhead compared to on-premises warehouses. Enterprises can choose from a variety of robust cloud data warehouses, including Amazon Redshift, Google BigQuery, Microsoft Azure Synapse, and Snowflake. Familiar tools and technologies used in the processing layer include PostgreSQL, Apache Spark, Redshift by Amazon, etc. (Oussous et al., 2018).

4.3.4 Data Visualization layer – deciphering the potential of big data

Finally, the visualization layer is the top-most layer in the BDA stack, where the actual analysis and insight generation materializes - the layer on which the end-users interact. Furthermore, the layer involves visualizations such as status reports, dashboards, and business intelligence (BI) systems. To these ends, data scientists and other technical users construct analytical models that empower businesses to understand their former operations and forecast what will happen and decide on how to improve the business going forward (Erraissi & Belangour, 2018). Hence, the visualization layer's components must make data simple to understand and manage. Going forward, this is often solved using visualization software such as Tableau, Looker, and Microsoft Power BI, which generate visualizations that allow users to make data-driven business decisions. Correspondingly, the author has exploited the Matplotlib library using Jupyter Notebook/Visual Studio Code as an integrated development environment (IDE) - a data visualization and graphical plotting library for Python, and a numerical extension to NumPy - to transform the ingested metrics into polar charts (Nelli, 2015). Moreover, the complete system (see Figure 11) integrating the holistic aggregation of each participating player chart into simple to understand visualizations is developed using Angular as a front-end framework - a platform for building single-page client applications (Angular, 2021).

4.4 Visualization method (Polar Chart)

Following the analytical stack's outcome, one obstacle with complex datasets is communicating the analysis findings intelligently to various stakeholders, which ultimately is necessary to support individual and organizational decision-making. Addressing this challenge, a disruptive science - that of data visualization - has advanced to examines how people rapidly can assimilate large amounts of information and generate visual representations of these complex datasets (Caban & Gotz, 2015).

Similarly, during the pre-case study, the author discovered the polar chart to be the most efficient way to benchmark and communicate a soccer player's ability. A representation baptized 'Rotelle' by the informants. Initially, the method or technique originated from the historical nurse Florence Nightingale, which utilized its functionalities to uncover sickness during WWII (O'Connor et al., 2020). By inventing the color statistical graphic entitled "*Diagram of the Causes of Mortality in the Army of the East,"* she dramatizes the degree of pointless fatalities in British military hospitals during the Crimean War (1954–56). This scientific collaboration enabled Florence to map the totals of death by month, with the area of each wedge representing the quantity of deaths which was further subdivided by colors founded on the cause of mortality. Hence, the chart illustrated in figure 13 uncovered that epidemic diseases were responsible for more British deaths in the course of the war than the presumed battlefield wounds – as the seminal diagram showed the majority of soldiers died from diseases shaded in grey than from wounds represented in red (O'Connor et al., 2020).

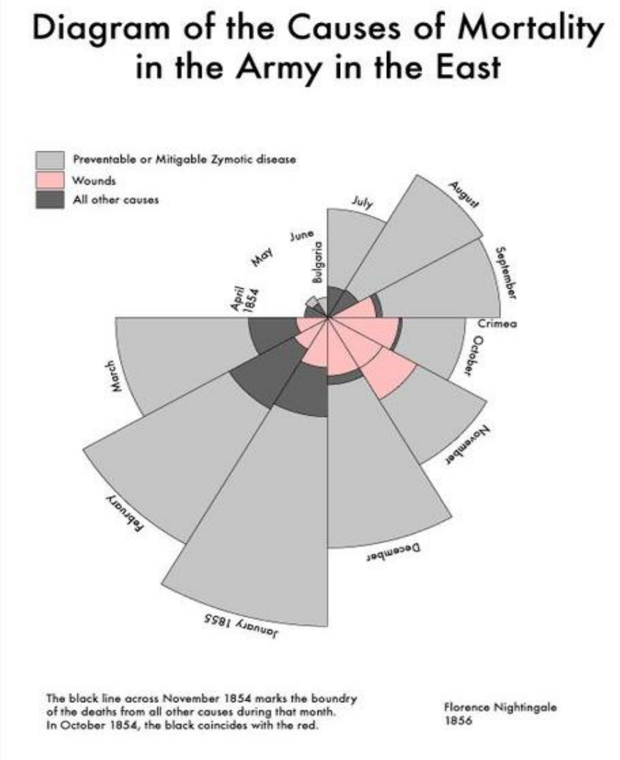


Figure 13: Florence coxcomb diagram on causes of mortality in the British army O`Connor et al. (2020).

Until recent time, this figure was generally referred to as a coxcomb or 'rose' diagram and enabled several comparisons of information to be seen in a single instantiation. However, for now, people refer to it as a polar chart. Additionally, it provides those unfamiliar with statistical data and reports with a helpful display (Magnello, 2012; Nelli, 2015). Furthermore, exploring the story behind how the chart got its titles and its possible use cases, Kirk (2016), the author of 'Visualizing Data,' labels it as:

> A polar chart shows values for three or more different quantitative measures in the same display. It uses a radial (circular) layout comprising several equal-angled sectors like slices of a pizza, one for each measure. In contrast to the radar chart (which uses position along a scale), the polar chart uses variation in the size of the sector areas to represent the quantitative values. It is, in essence, a radially plotted bar chart (Kirk, 2016).

In brief, the chart is a hybrid of a bar chart (exploiting length as a pre-attentive tool) and a pie chart (radial in nature, narrower at the center where segments increase towards the top, i.e., polar coordinates). In fact, like the pie chart, the degree of each sector provides percentage data represented by the classification with respect to the total. As for the bar chart, the circular extension is the numerical value of that category - which helps us comprehend the form of the chart as a means for representing data (Nelli, 2015).

Hence, the author considers the rendered polar chart or 'rotelle' is ideally suited for quick comparisons across proven metrics, as done in the artifact (figure 14). In addition, because the 'rotelle' uses length and width to represent the chosen data points, the eye tends to notice the lower and higher values more rapidly when compared across the different players (Kirk, 2016).

Similar to Florence Nightingale, a data-driven system based on polar charts are introduced as a provider of adjustable pre-constructed templates, processing raw data from StatsBomb to create personalized data visualizations identifying given prospects according to preferences in section 4.5. In addition, the polar chart also provides flexibility regarding tailoring visualizations concerning the player/team/issue being discussed for specific reports. For example, in the pre-case study, the major European soccer club needed a new position-specific visualization - as the feature needed was not currently/adequately illustrated by the system they had on hand. Luckily, the polar chart's flexibility made ease to access the new information, and new data integration went smoothly.

At last, the flexibility of the polar chart utilized within the organization's data-driven system was often emphasized as it made it easy to communicate valuable information efficiently between stakeholders. In short, the viz highlights a player's actual statistical output in a quick and easy glance, that being potential signings, academy players, or first-team players.

4.5 Demonstration of use - user scenario

After establishing the design theory justifying the design decisions in the sections above, a user scenario is presented to instruct a proof-of-concept. Initially, to demonstrate the artifact's utility, the template in figure 14 is altered with essential metrics for the up-and-coming Briton Patrick Bamford's position and role. Following the user scenario, the artifact is used as a tool to explain how a manager could approach a tactical decision concerning the Briton by utilizing big data as a data-driven decision-making system. At last, the complete version of the application is illustrated in figure 16 and the initial prototype in figure 15.

It is important to note that the chosen attributes are conceptualized and therefore collated by colors representing how each data point interrelates. For example, red represents both attacking on/off-ball events, while green illustrates on/off-ball defending.

Interpreting the polar chart (Figure 14)

As described in section 4.4, the polar chart comprises a number of wedges (slices) representing a proven metric. Each wedge's length corresponds to the selected player's percentile rank for that metric compared to the players in the same league and position. The percentile rank is the percentage of scores within a dataset equal to or lower than the score. This is reversed, such as 'turnovers.' In sum, a more extensive bar is always better.

Hence, when interpreting Patrick Bamford's performance overview (Figure 14) so far in the 20/21 season, one immediately notices the number of filled percentile wedges, indicating Bamford's outstanding ability to create expected goals per 90 minutes (xG). Another striking attribute describing Bamford's skillset is the number of touches he generates per 90 minutes in the opposition's box (touches in box). Intuitively, drawn from the knowledge above and Bamford being a tallish player (1.85 cm), one would most definitely conclude that he would be scoring the majority of his goal with a header from within the opposition box. In turn, as a manager, you would probably shape your tactics accordingly – blasting in high crosses aiming at his head.

Nevertheless, as a contradiction to this highly intuitive opinion based on Bamford's physics, his aerial percentile wedge just ranks him above the average striker in the air – winning merely 68 % of the aerials he contends (so far this season). Hence, being a manager, you probably would vast your most dangerous weapon by playing a tactic exclusively based on high crosses, as your striker's strength seems to lie as much along the ground as in the air.

Interpreting the Artifact (Figure 16)

Further, to enhance the beautiful game's tactical approach, the artifact integrates all involved player's costumed polar charts into a system. Such as shown in figure 16. Providing such a fast and holistic overview of all playing participant's particular skill set enable managers to effectively alter their game strategies according to the players at their disposal - exploiting any weakness found relevant within enemy lines.

A proof of the artifact's utility is anchored in its ability to assess Patrick Bamford's playstyle and identifying the logic behind his previous performances, as well as how to counter it. For example, looking further into the already stated goal-scorer, one initially acknowledges the high scores he has generated regarding touches in the opposition box, goal creation-actions, and being a frequent target of attempted passes. An indication that Leeds probably depends more on their striker than other teams. Taking these remarks into perspective and assessing them towards Bamford's generally low passing attributes, few touches/involvements before entering the opposition box, and his low shot creation-action, it is evident that Leeds United strategy does not include the striker to contribute much to build-up plays – before entering the final third. On the contrary, he probably operates as a so-called 'fox in the box' (most goals come from tap-ins). If we process this information further and add Bamford's somewhat impressive pressure stats, one should consider countering his abilities with defenders scoring high on stats reflecting good capabilities when pressured. Such as pressured passes, progressive passes (often a means to avoid press), turnovers, and a generally high combination of defensive stats to handles Bamford's superior off-ball movement and pressure style. Hence, as illustrated in the prototype (Figure 15), a reasonable nemesis for Bamford would be his current teammate Luke Ayling (current center-back for Leeds) – being among the top defensive players at all these stats except for turnovers.

Another example of the application's utility is identifying the high possibility for Leeds to attack through their left flank - as these players tend to have a higher attacking contribution than their right side (reflected in their red bars in figure 15). Further, to increase readability and usability, a hover effect makes each chart bigger when hovered over. Radio buttons make it possible to change formations, and a drag-and-drop feature makes the chart moveable. In sum, the artifact aims at putting the user in a better position than ever to outmaneuver the opposition tactically.

Possession For passing & ball progression	Transition For dribbling & positioning	Attacking For goal-scoring & chance-creation	Defending For defending
Pass Comp % The percentage of attempted passes that successfully reach a teammate	**Turnovers Per 90** Includes every time a player dissposses or miscontrol where the player loses the ball	**Goal Creation-Action (GCA) Per 90** The two offensive actions directly leading to a chance, such as passes, dribbles and drawing fouls	**Rate Adjusted Tackles Won %** The number of successful dribbles a player makes per 90 minute, adjusted for their success rate.
Progressive Passes Per 90 Completed passes that move the ball towards the opponents goal at least 10 yards - excludes passes from the defending 40 % of the pitch	**Succesful Dribbles %** Percentage a successful attempt at taking on a player and pass theme while retaining possesion	**Shot Creation-Action (SCA) Per 90** The two offensive actions directly leading to a chance to shot, such as passes, dribbles and drawing fouls	**Interceptions Per 90** The amount of secsessfully interseption per 90 minutes
Passes into 1/3 Per 90 Completed passes that enter the 1/3 of the pitch closest to the goal - not included set pieces	**Rate Adjusted Dribbles Per 90** The number of successful dribbles a player makes per 90 minutes, adjusted for their success rate	**xG (Expected Goals) Per 90** Expected goals per 90 minute included set pieces such as penalty kicks.	**Ball Recoveries Per 90** Every time a player regain possession of a loose ball per 90 minutes
Passes into Box Per 90 Completed passes that enter the penalty area of the pitch closest to the goal - not included set pieces	**Progressive Distance Per Carry** The number of yards a player carries the ball towards the opposition goal per carry the player complet in the average 90 minutes	**xA (Expected Assists) Per 90** Expected assist Per 90 included set pieces	**Pressure Regain %** Number of times the team win the ball within 5 secunds after the player applyed pressure to opposing player who is receiving, carrying or
Pressured Passes Per 90 Passes made while under pressure from opponent	**Touches in 1/3** The amount of times a player touch the ball in the final third towards the oppenent goal	**Non-Penalty xG + xA Per 90** Non-Penalty expected goal + expexted assist per 90 minute	**Dribbled Past** Number of times a player is dribbled past by an opposing player
	Touches in Box The amount of times a player touch the ball in the opponents penalty area	**Shots on Target %** Percentage of shot attempts that is on target	
	Pass Target Number of time a player was the target of an attempted pass	**Aerials Won %** Percentage of aerial attempts that is succesfully won	

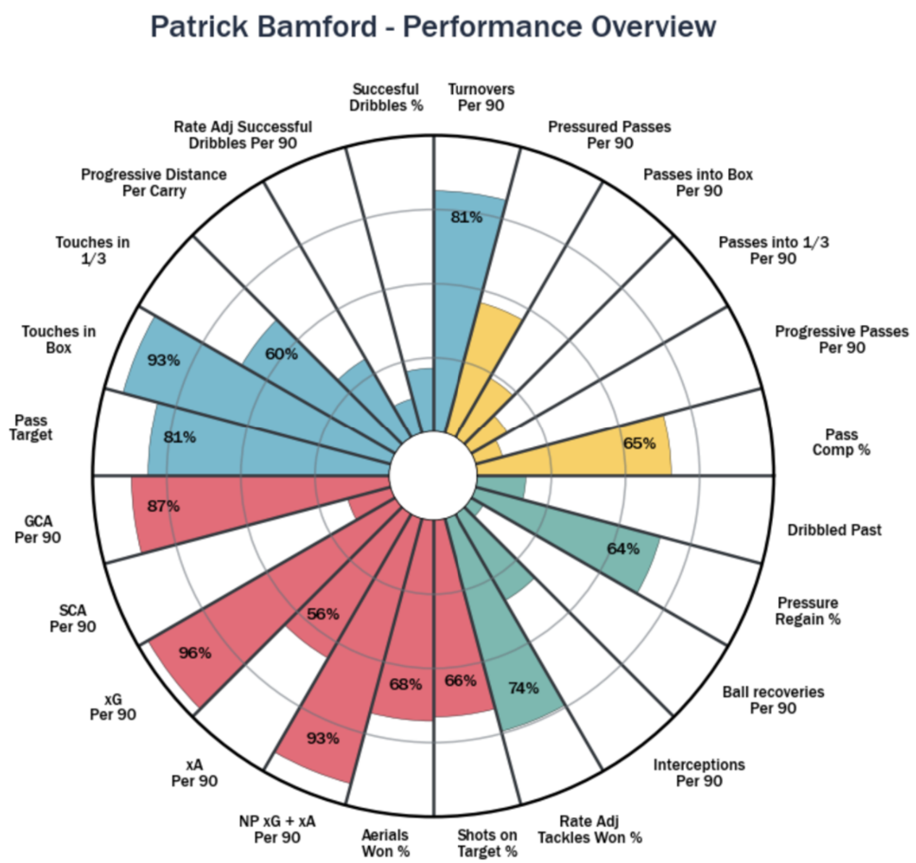


Figure 14: Interpreting the Polar Chart.

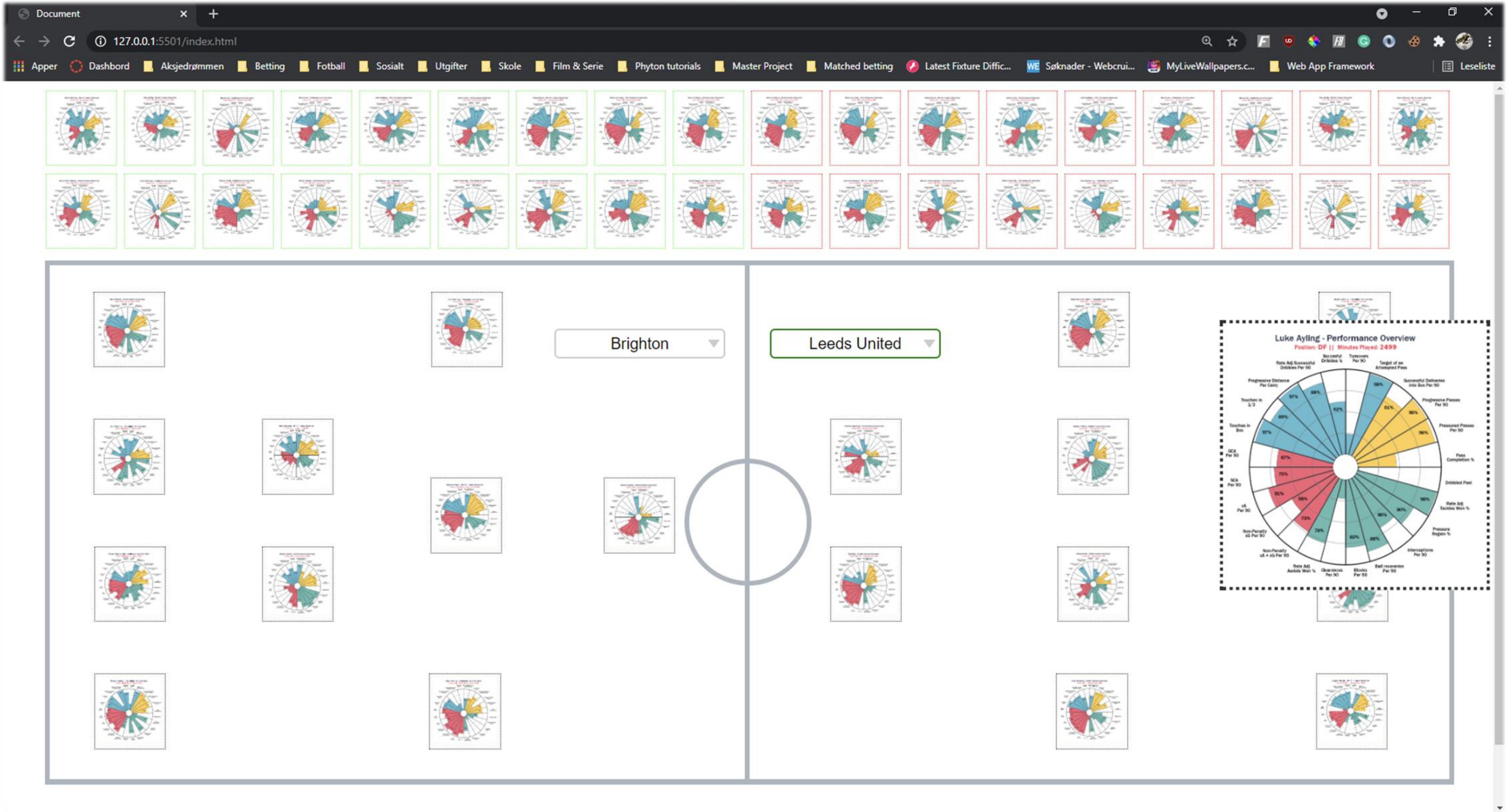


Figure 15: The prototype: left side and top-left side represents players at your disposal (names will be hidden due to experts may be biased when choosing between players they know, see Appendix A).

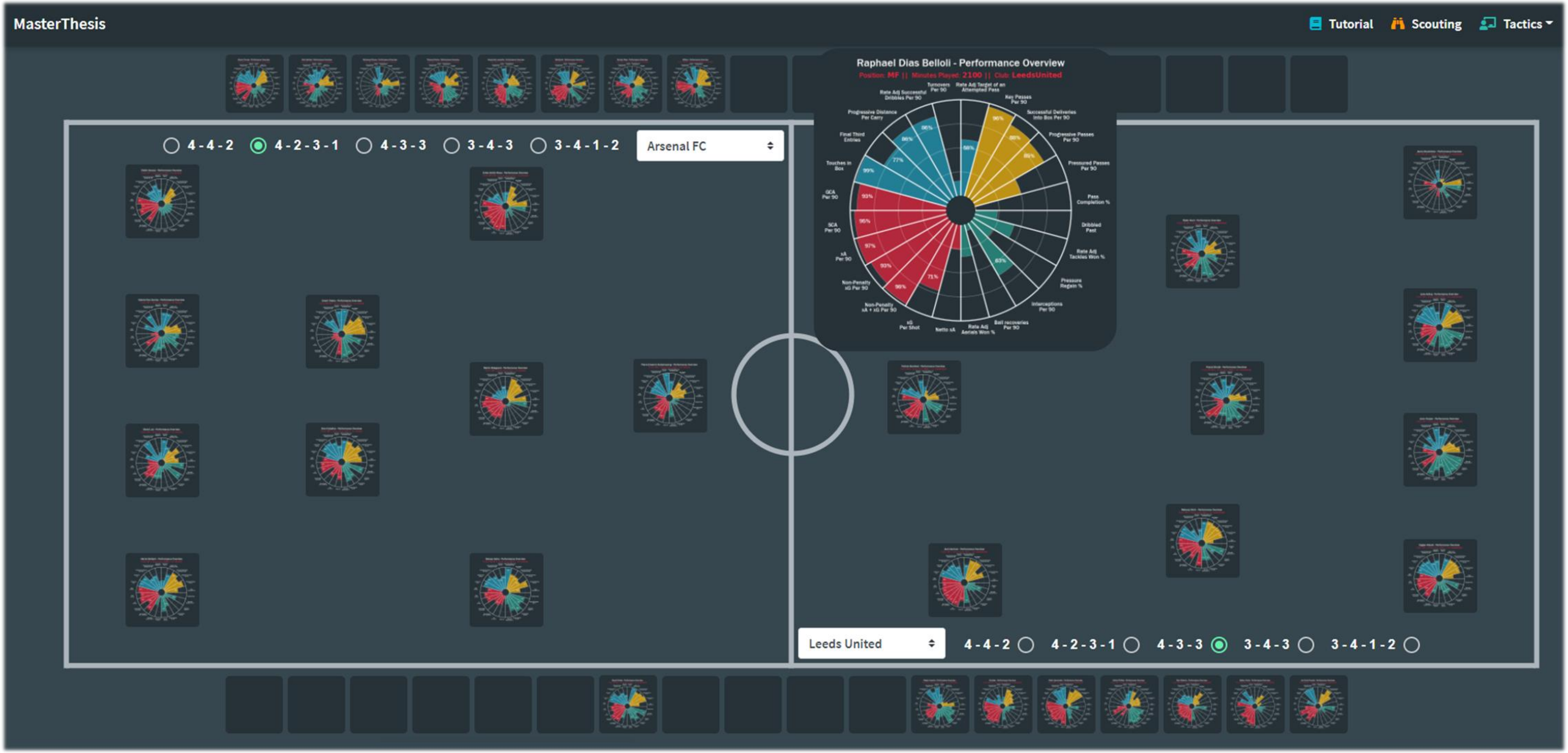


Figure 16: The proposed web application.

5. Findings: ex-post evaluation

In order to make honest statements and theorize about the artifact in use, the related prescriptive knowledge in section 4 is documented and collected in a way that accepts stepwise evaluations of the artifact as it emerges into this section's ex-post evaluation (see Section 3.3.2). Moreover, Sonnenberg and Vom Brocke (2012) claim that the exterior mode (Figure 5) mainly intends to analyze, describe, and predict what happens as artifacts exist and are used in their external environment. Additionally, the descriptive knowledge derived from the expert reviews is similar to a black-box test. Thus, the author can review significant design features to utilitarian ends and later compare the respondent's conventional approaches to the artifact in the discussion. Hence, to build consensus on the artifact's relevance, novelty, and importance in the chosen problem domain, this section manifests the results and findings (themes) extracted from the data analysis described in chapter 3.4.3.

Further, all respondents have been given anonymity to preserve the informant's privacy to the extent possible. In some cases, their role or/and title are mentioned without acknowledging their specific context. In addition, to ensure a certain language homogeneity throughout this thesis, the author has translated the respondent's direct quotes from Norwegian to English. Moreover, it is essential to note that the composed artifact is a component of a human-machine problem-solving system. For such developments, knowledge of empirical work and behavioral theories is crucial to construct and assess them. The constructs, components, and methods are therefore exercised within relevant environments by appropriate subjects. Because the proposed artifact represents the 'machine' part of the human-machine system constructing an information system, these sub-sections principal aim is to determine how well it works, not to theorize around or confirm anything concerning why it works (Gregor & Hevner, 2013).

5.1 Overview of central themes and consequential errors

Despite all the literary initiatives developed to standardize analytical practices in soccer, transcribed findings synthesized from the expert reviews suggest computer- and sport science current approaches causing more inefficiency across a soccer organization's manifold body than necessary. As researchers and scientists seek a revelation in new advancements, these actions are also the core extending the abovementioned issue, which not surprisingly originates from the *quantity of accessible tools.* Without the manifold of unproven tools, the respondents emphasized that the overall data-driven experiences in soccer would be less time-consuming. Furthermore, the amount of analytical content these tools generate longs for a comprehensive understanding between coaches and analysts. In Europe, these separate roles also tend to communicate in very different jargon, causing the analytical memo to halt across the organizational groups. Thus, even though a club inhabits a data-driven culture, all respondents emphasized that the most pivotal skill set to comprehend data value was mainly dependent on *two-way communication.*

Tied around these issues is the *utilization of analytical advancements.* In comparison, respondents have experienced various technological limitations interfering with their strategic decisions. For instance, one major restraint acknowledged during the reviews was the importance of valid data for the advancements to be *suitable for tactical decision-making*.

In contrast to the conventional tools, no respondent had any issue *applying their overarching tactical approach* to utilize the artifact. Simultaneously, their *reasoning for picking players* over others tends to be equivalent - both among themself and what happened for the club they managed in reality.

Moreover, when *comparing and validating* the artifact, the respondents emphasized their current approach to be less effective due to all tools available and the difficulty of choosing the right tool for the proper context.

Table 8 summarizes the main themes, sub-categories, and consequential errors identified during the data analysis. These findings are discussed in more detail throughout the rest of section 5.

Table 8: Overview of the data analysis main findings.

Main themes	Sub-categories & consequential errors
Data-driven culture	Two-way communication
The respondent's current approach to soccer tactics	Quantity of accessible tools
	Utilization of analytical advancements
	Suitability for tactical decision-making (valid data)
Utilization of the artifact	Applying a conventional strategy to the artifact
	The reasoning behind the player picks
Practical significance	Current approach vs. artifact
	Validation

5.2 Data-driven culture

The author finds this theme central as it captures the essence of lived experiences on the topic - and therefore - which capabilities the experts deem necessary for maintaining a sustainable data-driven culture. Correspondingly, the degree of analytical knowledge utilized within each respondent's environment also proves valuable in learning the artifact – as similar metrics tend to re-occur for those operating in a more data-driven environment. Furthermore, this has proven to provide a solid foundation for discussing their socio-technical behavior in granular details when grounding their artefactual decision-making. Hence, table 9 represents each respondent's historical background to get an in-depth understanding of how various roles with relevance to soccer tactics interact with soccer analytics – when discussing the matter in section 6.

Table 9: Respondent's background.

Respondent 1	Respondent 2
Relevant background & domain	
Data consultant:	**Performance analyst:**
"I have been a consultant for multiple large European soccer clubs, especially regarding the transfer part. However, considering the topic of this thesis being tactics, I also need to assess the players according to my client's tactical approach to identify a potential trade. Moreover, I have been around since the breakthrough of soccer metrics and applied all kinds of metrics in search of potential transfer targets. Unfortunately, in contrast to many others within the community, I do not have an educational background, as I came up the hard way."	*"First of, I am a practitioner on the subject as I started back in 2006 (with Interplay) and kept myself updated on new things from then. I have also taken numerous courses on the topic, including StatsBomb's. I would also add that the foundation of my experience includes all available UEFA-coach-licenses and an eternal passion for soccer."*

Respondent 3	Respondent 4
Relevant background & domain	
Sport science:	**Academy coach:**
"During my master's degree in sport science, I have utilized various analytical tools to measure multiple strength variables and analyze how training has affected relevant subjects. In combination and more relative to your task, I am passionate about soccer and athletics in general. I have also been informally involved in projects that contribute to the sports science part of soccer. Moreover, as the Norwegian soccer club, Start FK, has been engaged a lot in my study program, I am familiar with how various metrics are applied in soccer nowadays."	*"During my time at the university, I took the UEFA C-license, representing the first coaching license you get through UEFA. So, throughout these last couple of years, I have been practicing the license by coaching two men's academy teams and one women's academy team. One of these teams was Notodden (G13-G14) academy, which first-team played in the Norwegian second division, 'OBOS-ligaen,' at that moment. So, my current experience is similar to what the 'common man' views as conventional analytics – using whiteboards and some simple training applications."*

5.2.1 Two-way communication

The various ways of how people use metadata are commonly known as an issue within the analytics community. Which often leads to misinterpretation for an end-users desired contextual knowledge. There is always a subjective judgment in soccer whether the analytical information provided aligns with the coaching staff's strategy or what the data analytics deem essential. Moreover, there is one particular factor affecting the consensus stated above, as all respondents emphasized the importance of using a universal jargon, as expressed by the freelance data consultant:

> "*From my perspective, there are two kinds of people. First, we have the people working specifically with soccer relative stuff - such as the coaching staff and the club recruiters who typically apply conventional approaches to their work. E.g., they travel to scout players instead of identifying potential players with data first. Second, we have the 'nerds' or data analytics which apply newer, more innovative methods. Hence, in my line of work, I will define one of the most considerable skills making the communication between these two sides more efficient. Unfortunately, as of now, they are not talking the same language. That typically means contextualizing a player's ability shown in the output or contextualizing why some data do not always speak the truth you seek.*
>
> *Another exciting aspect is the disconnect in communication between head coaches and scientists as most elite trainers are very conventional and hereafter do not trust data – e.g., how many coaches do you see under 50-60 years? There are obviously some managers over 50 that also overuse analytics - I once heard a rumor that Bielsa had a 20-page notebook on a GK."* (Respondent1)

Another respondent extends to the argument stated above and implicitly refers to communication as a two-way skill. In addition, the author also learned that little research focuses on applying knowledge management in connection with user support. Moreover, one needs to understand other people's perception of strategy to present evidence making a desirable or positive outcome for all stakeholders:

"As an analyst, my job is to interpret, parse and communicate the contextual side of the data we gather so that anyone can grasp it. However, to make this process sustainable in a soccer context, the most valuable skill set is always to comprehend the head coach's perspective on game strategy and game model. Moreover, to do so, you need to treat these perceptions with a bureaucratic and pedagogic approach." (Respondent 2)

5.3 The respondent's current approach to soccer tactics

This theme establishes the respondent's current approach to soccer tactics and the difficulties of applying proper technological advancements to support tactical decision-making in their current practice.

5.3.1 Quantity of accessible tools

Despite all the literary initiatives developed to standardize analytical practices in soccer, the expert reviews expressed computer- and sport science current approaches causing more inefficiency across a soccer organization's manifold body than necessary. As the organizations seek a cutting edge through new advancements, these actions are also the core extending the abovementioned issue, which not surprisingly originates from the quantity of accessible tools. The table below summarizes some of the leading providers and technologies that re-occurred during the interviews.

Table 10: Summary of technological tools.

Tool	Description
Wyscout	Wyscout is an Italian company that supports soccer scouting, match analysis, and transfer dynamics (Wyscout, 2020).
InStat	InStat is a sports performance analysis company providing professional tools for individual and team performance evaluation and scouting (Instatsport, 2020).
StatsBomb	StatsBomb is a company providing a brand new proprietary dataset with granular data for powerful analytics for various sports (StatsBomb, 2020)
Playermaker	Playermaker is an intelligent motion sensor attached directly to a player's boot, producing over 30 performance metrics (Playermaker, 2020).
Veo	Veo soccer camera is a complete solution for soccer recording, coaching, and analysis, as the AI-powered camera can record without a cameraman (Veo, 2020).
GPS and LPS	GPS and LPS tracking let soccer cubs gather biological data for injury prevention, training load, and overall improvement (Hennessy & Jeffreys, 2018)

Addressing all these available tool's effects on tactical decision-making, most respondents agreed that the present quantity of tools caused inefficiency - primarily because they provide very similar functions:

"*There is too much work scanning through all these similar tools we have at hand. Especially concerning the time. I feel we are in a position where we need to aggregate the ones that will be proven over time."* (Respondent 2)

5.3.2 Utilization of analytical advancements

With the present quantity of tools in mind, the data consultant mentioned data analytics to be almost synonym with player recruitment and identifying new talents. A statement that is drawn from the fact that machines probably possesses the capability of outworking a traditional scout's former tasks:

"*Nowadays, data analytics is almost synonym with player recruitment and identifying new talents. However, there are absolutely some valuable metrics for the club's match-to-match use, as most match-to-match tactical analysis also depends on identifying the oppositions playing patterns. Another critical factor here is the limitation of the human eye – you cannot watch tape 24/7 minus sleeping hours. In contrast, you can parse and locate these key factors in only seconds with machines. Thus, you can, e.g., verify that the patterns you intuitively have recognized about your rival as accurate. Based on this, I would most definitely say there is a massive advantage in locating these patterns through data as it speeds up the process."* (Respondent 1)

As indicated in the direct quotation above, there are acknowledge metrics for match-to-match use - as most tactical analysis depends on identifying the oppositions playing patterns based on the same metrics as utilized in their player recruitment. Another critical factor mentioned above is the limitation of the human eye – as you cannot watch tape 24/7 minus sleeping hours. In contrast, machines are used to parse and locate such vital factors in a matter of seconds, as expressed by the respondent. In addition, the same respondent reflects on how he currently uses the analytical advancements for this purpose:

*"Considering my approach to tactical analysis, I often look at team-based stats such as PPDA (passes per defensive action) and vice versa. Further, I assess the average player position during a game, where chances are created, and what context they root from. For now, video is the primary tool. However, we also use player-centric stats. For example, suppose the opposition tends to use a formation with two strikers. In that case, I will look at their metrics to identify patterns and exploit this knowledge to predict their collaborative tasks and counter it." (*Respondent 1)

Moreover, similar patterns, as stated above, are often integrated into match reports and presented to the player groups. To these ends, they can practice towards what the staff predicts will happen in an 11 vs. 11 before game-day, as expressed by respondent 3:

"*If we were to look at the tactical aspect, we utilize metrics quantifying team performances, injuries, player performance, etc. A typical example would be to locate the opposition's most vulnerable position and attack him with our most skilled player. For example, if you have a player like Neymar, you will probably try to isolate him against a slow opponent. In this example, the opponent would obviously be ready for such a thing. Still, we need to figure out a strategy preventing the opposition from defending what they anticipate us to do."* (Respondent 3)

Further, as the respondents elaborated on the objectiveness of data, three of the respondents agreed on what happens when data contradicts their perceived intuition, as stated in this quote:

"*From my perspective, I think that when data contradicts my personal opinions, I'm more interested in finding out why the data says otherwise than finding flaws in my intuition. And, as I stated before, data is objective, and there are limits to how many games I can watch. Based on that, a player can either have been outstanding in the one match I saw and wrongly affected my intuition as he could've been bad in the rest of his games. So, I would almost always trust the data - but then it all comes down to how the data we use is contextualized. Take Lewandowski at Bayern as an example. Even though he misses the most chances, this does not necessarily make him bad, as the data also* *shows him to be the one who creates the most chances. The same goes for what I deem as a contextual flaw in Arron wan-Bissaka's dribbled past stat, reflecting him to be bad in one-on-one situations as players often go past him before he wins the ball back. In accordance, you can ask yourself the question; have you really been dribbled past if you win the ball back during the same sequence? In sum, you need to assess stats towards each other to get an understanding of what actually happens."* (Respondent 1)

Nevertheless, as three of the respondents perceived the topic as stated above, one respondent mentioned he never been in a situation where the data has affected his decisions in a negative manner, other than apparent bugs in the system - as they always quality check the data themselves:

"That is a tricky question. I do not believe I ever have approached a problem like this - but say, if we have data telling our opponent to play 4-4-2, but during the game, we observe that they are playing a 4-3-3, then my intuition would overtake the data. For me, when exploring these types of challenges, we fall under the terms scouting and opponent analysis. A good example is our own fullback, who scores high on winning duels in the statistics. Still, when he loses some duels, these are often vast mistakes leading to chances, which is not shown in the statistics yet. Hence, when we work with all these different tools, there is always essential to quality check what we generate, clean the data appropriate for our KPIs and player modules, and organize weekly plans for when and on whom the different data should be applied." (Respondent 2)

5.3.3 Suitability for tactical decision-making

Even though all respondents agree that it is money thrown out the window not to have the analytics perspective, a common issue is how to acquire the proper tool among all tools available. For clubs that do not yet participate, the resistance is often in the top management. The typical structure consists of an old guard, often resistant to tech, and a pretty young data team trying to break through. Hence, the cost is not an issue; it is more about what works for each group. Furthermore, there is a consensus among the respondents that clubs usually acquire tools deemed the most sophisticated tools on their level, not which one is most suitable for tactical decision-making. One of the respondents explained the process of acquiring tools for physical attributes as follows:

"Tools we buy and acquire are often proprietary software proven to be reliable in other domains. So, a criterion is that they are established and measure what we actually try to test. For example, we use MuscleLab as our leading software to measure and gather biological data, a very dominant software in the community. For analyzing such results, we can use a common tool called SPS." (Respondent 3)

In contrast, the academy coach explained in short detail how this often varies for youth academies with fewer resources, while the performance analyst stated how the need for customization trumped in the choice of more in-depth software:

"We use these tools as they are offered to the club on a discount, and the club preferred to use its resources in other places." (Respondent 4)

"When talking about our own KPI, this was the reason for using Interplay over all these years. However, when Wyscout (huddle) approached the scene, we thought their KPIs to be easier to customize to our needs." (Respondent 2)

5.4 Utilization of the artifact

After elaborating on their current methods, the respondents assessed/tested the utility of the web application illustrated in Appendix B, according to the use case presented in section 3.4.2 and Appendix A. Hence, they were asked to apply their regular game philosophy when choosing a desirable line-up, pretending to be the active manager of Tottenham Hotspurs. In order to exclude expert bias, the respondents approached the assessment without any knowledge of which team to managed or which to face. As a result, they objectively selected which players to include in their strategy against a very destructive Burnley-side.

5.4.1 Applying a conventional strategy to the artifact (strategy)

As the respondents followed the abovementioned use case, they had no problem integrating an individualized strategy when choosing a line-up. However, as expressed below, the respondents tend to utilize two distinct tactical approaches. First, the perhaps most tactical qualified respondent - the performance analyst - located weaknesses within his opposition before choosing which player to exploit his findings, as quoted below:

> *"So, the most important thing to start with is which channels to attack in their defensive line. By this, I mean how dynamic are their four defensive players compared to each other, per strengths, weaknesses, threats, and opportunities. The exact process goes for both the MF and DF. After a fast glance, I think it is imperative to include players with a majority of high attacking attributes for this particular game - as the visualized statistics indicate that I not only face one low block but two."* (Respondent 2)

Second, in contrast to the performance analyst, the rest of the respondents identified which player they found most suitable for each position according to their philosophy. Hereafter, they scanned the opposition for weaknesses and strengths and rotated some positions to increase pressure on the opponent's weakest links:

> *"I would first locate the players with the best aggregation of which stats I deem necessary in each position. Then I would assess these stats against the direct player they face - meaning, my best dribbler should execute his work versus their weakest defender and vice versa."* (Respondent 4)

Further, as the respondents read the demonstration of use (see Appendix C) before the test, one of the respondents elaborated on how he utilized the polar chart to identify what he deems necessary according to his ideas and how much trust he could lay in the feet of each player:

> *"For these positions, I value defenders with an overall high ball-control, which is reflected in the blue bars of the chart, and obviously the defending stats illustrated in the green bars. Further, I think playing time is of the essence, as it probably displays how trusted these center-backs of choice are."* (Respondent 3)

Another respondent also enlightens how he compared polar charts in order to judge players co-existence or chemistry on the pitch by adding minutes played with positional information and how two players statistically could complement each other:

> "As I a*ssess these center-backs, I always go for the players with good chemistry in these positions. Many minutes for the two players of choice indicate that they have played together a lot. Their defensive bars also indicate that the players have clear roles - as*

one being more aggressive, has higher interceptions, and is stronger in the air than the other, which tends to block more and tackle as I anticipate him to secure his teammate, playing like a second defender." (Respondent 4)

5.4.2 Reasoning behind the strategic decisions

Correspondent to their philosophy, the respondents fancied their strategic decisions on a typically possession-oriented style of play as the opposition defended with a low line. In accordance, the experts consistently chose the players with the highest attacking and transition score in almost every position, as shown in table 12.

Table 11: The respondent's tactical groundings.

Position	Finding	Grounding
Center backs	For the center-back positions, the respondents focused on overall high defensive attributes and the ability to handle the ball under pressure. As the experts were assigned with a formation consisting of two center-backs, there was an intriguing awareness of locating defenders with complementary capabilities. As a result, a potential logic for identifying chemistry was proposed.	"As I *assess these center-backs, I always go for the players with good chemistry in these positions. Many minutes for the two players of choice indicate that they have played together a lot. Their defensive bars also indicate that the players have clear roles - as one being more aggressive, has higher interceptions, and is stronger in the air than the other, which tends to block more and tackle as I anticipate him to secure his teammate, playing like a second defender."* (Respondent 4)
Full backs	For the full-backs, the respondents desired these players to have the most dynamic aggregation of high stats. Moreover, as all experts anticipated a modest pressure from the opposition, they required these 'attacking' defenders to recycle the ball and create chances. To this end, the two most versatile players with the highest attacking output were favored by all the experts, as shown in table 12.	*"So, I would look for the defensive players with the overall best attacking- and ball retention attributes for the full-backs positions. These are natural stats to select for any player in a fullback position. These choices are also complementary as the opposition team seems to play with a shallow block. In this case, I would need creative players in these two positions to break through and use the possession I expect to have for this game."* (Respondent 1)
Defensive Midfielder	All respondents valued a defensive player to be the tissue connecting the team. Accordingly, 4 of 4 respondents chose whom they deemed the best defensive midfielder towards the left, as he should also be the caretaker of the opposition's most creative midfielder. As a result, 3 of 4 respondents chose the same player.	*"This midfielder (P. Højbjerg) seems to be one of the best 'Backbone' players in the league. He also contributes with a very dynamic skill set, such as high accuracy on various passing types, and he rarely loses the ball. In my opinion, the best choice for keeping the team in balance. As I see this position's opposing player being their best central midfielder, I would play my best defending player here, and vice versa for my RCM."* (Respondent 3)
Central Midfielder	Like the defensive midfielder, the central midfielder`s polar chart should entail a versatile player. However, to	*"I look for a more dynamic and offensively good player for this position. This player (T. Ndombele) could really take advantage and*

	exploit what the respondents found to be the opposition's biggest weakness – their left side and primarily their central left midfielder – they all desired a player with a more creative spirit in this position. As a result, 3 of 4 respondents chose the same player.	*make things happen where the opposition seems weakest - this being on my right side. Putting the most overall player on this side would make him face a lousy player who almost always would be the recipe to create an opening against a low block."* (Respondent 4)
Wingers	The key aspect discussed for choosing a winger was the opposing full-back's tendency of getting tricked - in terms of their *dribbled past* attribute being low. As an outcome, the respondents scrutinized their troop for the attacking midfielders with the best mixture of transition stats to take on the opposition's weak spots. Equally important, the experts acknowledge their competitor's left-oriented players to struggle more as their most creative player (Dwight McNeil) leaves behind more space around his starting position in order for him to 'spill his magic.'	"*On the flanks, I always look for the most capable transitioning players. Especially as the two-opponent fullbacks tend to be weak in one-on-ones, as well as mediocre tacklers, overall, these players at my disposal are quite spectacular when it comes to offensive contributions and creativity. He* (*Son Heung-min*) *may not be the best dribbler, but he is well over average, and the best dribblers also tend to be the ones to fail the most as they also are the players who dribble the most. As I view this guy as my best winger, I would place him on the right side, as my opposition seems very weak defensively on this side."* (Respondent 3)
Central Attacking Midfielder	There was an overall consensus stating the central attacking midfielder to possess high creation and progression stats, as the wingers should be the ones to dribble. However, there seemed to be a variation in which attributes the respondents deemed most pivotal within these two aspects. For example, one respondent used the same arguments for his choice but ended up with a pretty different player (Dele Alli) than the others (Harry Kane and Gareth Bale).	*"So, this player`s (Harry Kane) stats are really something. It`s quite unique to be the most creative and dangerous player at the same time, which immediately tells me he needs to be the center of attention."* (Respondent 1) *"This must be the new Frank Lampard. High attacking contribution and some important progression stats. Considering his mixture of attributes, he* may *not partake enough in build-ups. However, he probably possesses excellent movements into the box – which could be very valuable in opening the opposition`s fortress."* (Respondent 3)
Striker	At last, all respondents explicitly focused on an overall high combination of red bars, and especially xG was the center of attention. The chance of recycling the ball through pressure was also of the essence for one of the respondents.	"*On top, I may choose a player with good defensive capabilities, as both CB*'*s seems to play almost every pass under pressure, but at the same time, they have a pretty low accuracy to their passes as they seem to often clear the ball. So I expect to recycle the ball quite often and need both an excellent pressing striker and a creative striker to break down the block. I would also go for the highest xG on this player as he should be my main treat."* (Respondent 1)

5.5 Practical significance

This section confirms how the respondents compare their conventional approach to tactics against their experience with the artifact. It also validates the final line-up for each respondent against what Mourinho chose for his Tottenham against Burnley.

5.5.1 Current approach vs. artifact

None of the respondents had seen a similar construction to the web application in their current work. They agreed on the artifact to be unique and more practical than what they have used so far, as expressed below:

> *"From my point of view, this method would absolutely improve the current approach to tactics. For example, if you are in a position where you can impact the line-up of your team, you would probably have a reasonable understanding of your own team's dynamics. Still, with this artifact, you can scan your opposition efficiently through the colors representing their attributes in different parts of the game and then assess how to take advantage of them. Based on all the various data tools we possess today, I assume this would be a much more efficient way of approaching tactics in the future.*
>
> *E.g., during the use case, I could pretty fast locate my players to penetrate their left side - as their left central midfielder seemed very weak, and their left fullback is pretty bad in one-on-one situations. We can also see that their best-attacking midfielder plays on the left side, which would probably leave these two weak players vulnerable in transition."* (Respondent 1)

One of the respondents also mentioned that if he could alter all features according to his philosophy and add some stats, this could very well be something he pays for:

> *"Yes, short and good. I would think the construction work could be problematic if we had to do it ourselves. Still, if I could have it as you have shown me, I think it would be advantageous. Something that surprised me is that I wonder if the hiding of names actually should be a feature in itself. As names sometimes function as noise, cause some trainers often lay their love on players, even when they do not perform. It could be interesting to have a dropdown menu where you could select stats for various situations, such as players xG on corners or free kicks, etc."* (Respondent 2)

Further, the less tactical respondent (sport science) emphasized the importance of customizing the charts on the user's premises. In addition, he also mentioned the easiness of learning the artifact as the visuals were quite appealing and well documented:

> *"Absolutely. As we talked about, a coach would obviously know his team. Still, with this tool, a specific stat the coach was not aware of could occur and change the line-up for a given rival. I do not possess the experience to address all the metrics properly for now, but I think I would have learned it with a day or two, as the logic behind the system and the visuals are quite easy to understand. Correspondingly, I assume this tool would be very beneficial for a coach who knows his stuff and likes to dive into the material. When it comes down to stats, every individual view different stats as essential and not, so it would be crucial to make the charts on the user's premises."* (Respondent 3)

At last, the academy coach mentioned the educational benefits of utilizing such a tool. The potential outcome of absorbing critical information concerning how players operate within a system could also help a player's - less measurable - tactical ability for games to come. As an extension to players gaining more holistic perspectives, this could also enhance their relations with coaches:

> *"Yes, I think so. It would have brought me closer to the small details that all managers seek to improve their game. I would really have gone into the depth of this approach if I could - cause as a manager, you have this ultimate desire to win at almost all costs, and if I feel this is something that could help me achieve that, I use it. Just consider how the potential result of players absorbing critical information concerning how other players operate within a system could improve their tactical ability. As an extension to players gaining more perspective on the game, this could also help coaches and players to understand each other on a whole new level."* (Respondent 4)

5.5.2 Comparison and validation of line-ups

In order to validate the artifact further, all the respondent's final line-ups were compared to the actual line-up of Jose Mourinho. Of course, as all people have different philosophies and preferences, this is not a significant validation. However, it was fascinating to observe how all the respondents, being very possession-oriented and creative in their philosophy, in contrast to Mourinho – commonly known as a defensive strategist - ended up with almost the same line-up in approximately 20-30 minutes.

Table 12: Final line-ups compared to reality.

Pos.	**Data consultant**	**Performance analyst**	**Sport science**	**Academy coach**	**J. Mourinho**
LB	S. Reguilón	S. Reguilón	S. Reguilón	S. Reguilón	**Sergio Reguilón**
LCB	Eric Dier	T. Alderweirled	Eric Dier	Eric Dier	**T. Alderweirled**
RCB	D. Sanches	D. Sanchez	T. Alderweirled	Ben Davies	**D. Sanches**
RB	Serge Aurier	Serge Aurier	Serge Aurier	Serge Aurier	**Serge Aurier**
LCM	P. Højbjerg	P. Højbjerg	P. Højbjerg	Moussa Sissoko	**P. Højbjerg**
RCM	T. Ndombele	T. Ndombele	T. Ndombele	P. Højbjerg	**T. Ndombele**
CAM	Harry Kane	Gareth Bale	Harry Kane	Dele Alli	**Lucas Moura**
RW	Erik Lamela	Lucas Moura	Son Heung-min	Gareth Bale	**Gareth Bale**
LW	Son Heung-min	Harry Kane	Erik Lamela	Harry Kane	**Son Heung-min**
ST	Dele Alli	Son Heung-min	Gareth Bale	Son Heung-min	**Harry Kane**
SUM	**8/11**	**11/11**	**9/11**	**7/11**	

6. Discussion

Throughout this thesis, it is evident that the statistical revolution of data-driven decision-making has established itself as the nexus navigating modern soccer towards the future. To this end, the author has ascertained how related works have approached the relatively untouched domain head-on - providing a comprehensive array of game-changing metrics. However, as the director of data science at STATS states in a documentary called *the numbers game*, it is not about gathering the data, as it already exists - it is about creating a universal language of that data (Lucey, 2017). Therefore, when interpreting the initiated research question drawn from the thesis problem definition (see Section 1.2 and 1.3):

"*How can player-centric metrics be utilized to simplify tactical decision-making in soccer?"*

This thesis aims to assist managers, analysts, and omnifarious soccer personalities in recognizing a more streamlined process of consuming tactical knowledge. Moreover, to address the issues associated with bridging the abovementioned research gap, the author has constructed a knowledge contribution in the form of a novel web application. To this end, the main target of the discussion is to compare the theoretical foundation forming the system against the current methodological culture of soccer experts to analyze the web application's *utility* and ultimately highlight the artifact's *practical significance*.

6.1 Theoretical implications

Equivalent to how Lucey (2017) argues for transitioning all existing soccer data into a universal language, the web application intends to partake in a comparable conversion by extending to soccer's traditional whiteboard. Moreover, to illustrate each playing participants contribution both individually and in collaboration, player-centric performance data like the sensitivity attributes extracted by Hassan et al. (2020), the omnipresent expected values discussed by Kharrat et al. (2020), and the off-ball metrics conducted by Llana et al. (2020) were aggregated into polar charts reflecting each player's capabilities. In addition, it became evident during the expert reviews that similar KPIs were a rather pivotal part of the respondent's current methods. However, as expressed by most respondents concerning the current tool's suitability for tactical decision-making, there is too much work scanning through all available means. Therefore, it is reasonable to argue that these massive amounts of various data tools have caused a damper for Lucey's (2017) universal language.

Nevertheless, a consensus stating the artifact to generalize the strategical assessment of power ratios between two opponents was in terms of its time efficiency established among the respondents. Moreover, this finding emphasizes how related works currently impact conventional methods, which ultimately strengthen the artifact's utility according to Spearman's (2018), McHale, and Relton's (2018) statement on the necessity of these performance indicators to be of use to the end-user.

Another issue extending to the necessity of contextualizing performance indicators on the end user's premises is the metrics contradicting the respondent's personal opinions. A re-occurring example during the interviews was the discussion on the metric *dribbled past*, a defending metric classified by StatsBomb in the glossary as *"number of times a player is dribbled past by an opposing player."* Although, in this case, the metric is meant to reflect a player's defensive one-on-one capabilities, it tends to award highly regarded defensive players with a low score against other defenders. Thus, as an example, the former Palace full-back Aaron Wan-Bissaka, reckoned as a master in the dying art of tackles, scored negatively on the dribbled past metric -

meaning one of Premier League's most feared tacklers, often to be dribbled past by an opposing player. In this scenario, one respondent highlighted that we need to realize when to look beyond superficial numbers. Hence, when reviewing Wan-Bissaka's style of play by utilizing the polar chart illustrated in figure 17, he argued the metric to count each time an opposition player moved the ball past him as a dribble, even though he won the ball with a late tackle in the same sequence. As Wan-Bissaka output also showed him to have among the highest rate adjusted tackles won for defenders. In accordance, another respondent stated this to be a contextual flaw in the data, asking the question, *"have you really been dribbled past if you win the ball back during the same sequence?"*

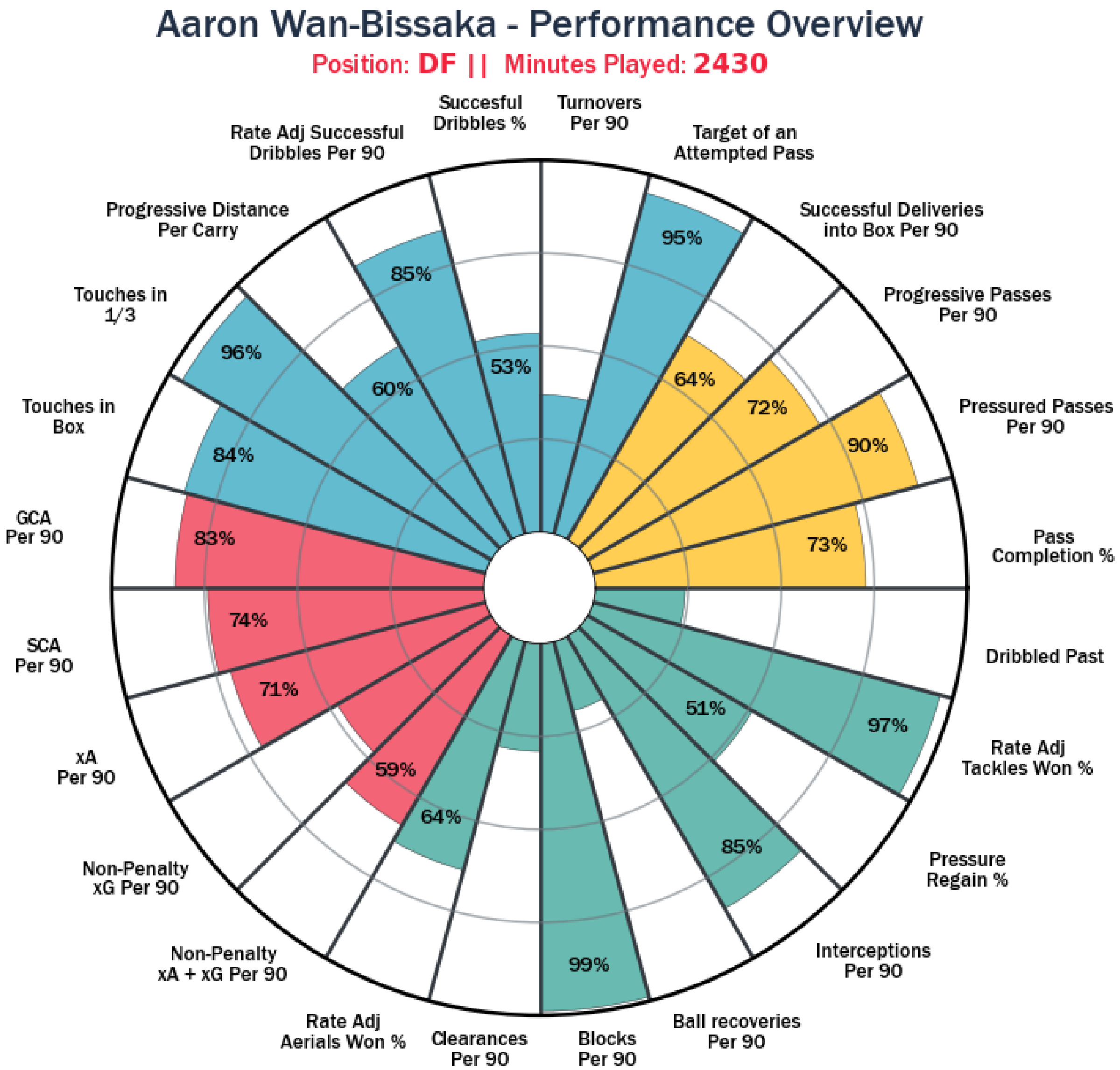


Figure 17: Aaron Wan-Bissaka - performance overview

Thus, in retrospect, it can be argued that the web application provides a holistic overview enabling managers to evaluate the inferred statistics as interpretable synergies, preventing a variety of stats from the crucial limitation Decroos et al. (2019) refer to as 'rigid values' - or lack of context (see Section 4.3.2).

Furthermore, as the respondents utilized the web application in corroboration to their strategy, most respondents located similar patterns to identify both player chemistry and individuals play of style. In accordance, the perhaps most intriguing outcome and prominent finding were how one of the respondents reasoned for his choice of center-backs, as he assessed player chemistry according to what he deemed as clear roles. As a result, by allocating the dynamic interplay of various player's polar charts into categories consisting of cohorts with complementary capabilities, the respondent implicitly proposed a potential logic for identifying chemistry among players. In turn, the artifact could potentially thrive in combination with Dick and Brefeld's (2019) and Van Haaren's (2020) research, as the researchers examined players' positional context and chemistry to reveal cutting-edge synergies within a group of players.

In summary, these elaborations amplify Kharrat et al.'s (2020) considerations on traditional performance indicators simplicity to some extent - as these metrics underlying algorithms lack context and a deeper understanding of the situations in which actions are committed. While this is the case, when symbiotic information is provided separately from the different analytical advancements presented in section 5.3.1, the respondents found the systematic visualization of the artifact to present a deeper understanding of how the color-coded aggregation of position-based stats revealed critical patterns. For example, as the performance analyst applied his traditional approach to strategically assess the opposition using the artifact, he emphasized the advantage of how efficiently he found himself to identify which channels to attack in their defensive line – compared to prior experiences. In contrast to the traditional performance indicators, the artifact amplified the holistic understanding of how coaches efficiently could exploit this knowledge to their advantage. At the same time, the combination of aggregating positional player-centric stats (data points suitable for a player's default position) in a system with familiar features, the artifact reduced the time-consuming process of scanning through various software.

Finally, one particular factor affecting the consensus stated above, as all respondents emphasized the importance of using universal jargon, is how the artifact tends to be capable of telling the same story to both a coach and an analyst. A statement that is strengthened as one respondent, working in the sports science domain, expressed:

> I do not possess the experience to address all the metrics properly for now, but I think I would have learned it with a day or two, as the logic behind the system and the visuals are pretty easy to understand (Respondent 3).

Ultimately, this strengthens the web applications proof-of-use as Decroos, Bransen, Van Haaren, and Davis (2019) elaborate on the importance of structuring a system according to the context it is supposed to enhance.

6.2 Practical implications

As with any contribution of this type, the sole intention of the proposed web application is to serve as a blueprint for future work. As a result, researchers can practice their discretion to vary what is proposed or submit and achieve improvements (Gregor & Hevner, 2013). In accordance, the application is mentioned by the respondents to show a promising potential if configured on the stakeholder's premises. In order to discuss the significance of the application in light of practical contributions, the author finds it valuable to compare the findings towards the existential reality of the Tottenham versus Burnley match. This allows contrasts to be drawn amid instantiations of artifacts and abstract knowledge and grants subjective impressions and experiences from designers and respondents.

To these ends, an intriguing re-occurrence of practical significance is the resulting consensus among the experts to either attack Burnley along their left flank or their left midfielder.

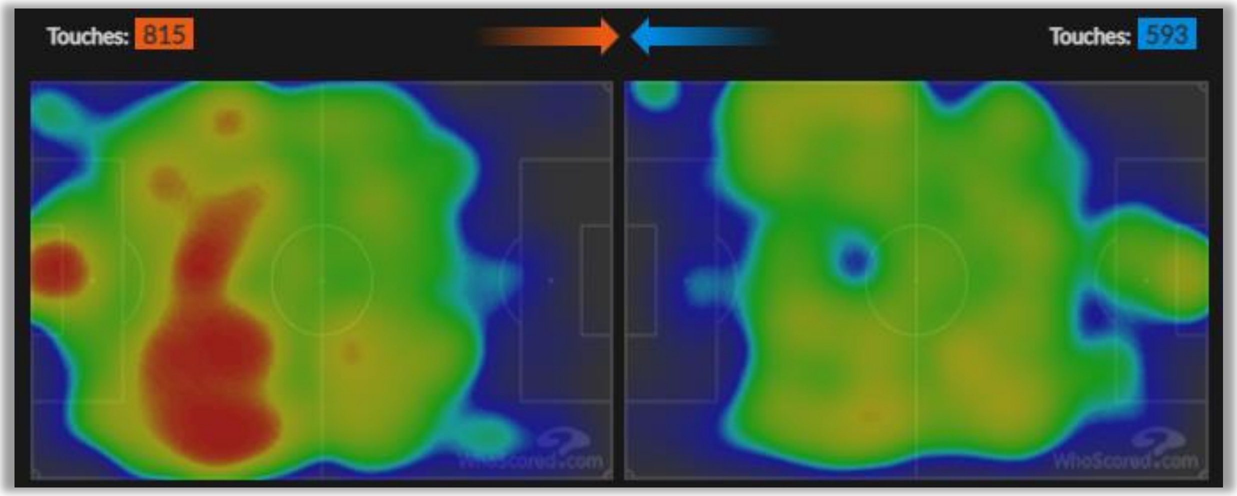


Figure 18: Heatmaps retrieved from whoscored.com (2021).

In accordance, the heatmaps retrieved from whoscored.com (2021) reflect how the majority of Tottenham's 815 touches were centralized at those exact locations - resulting in a 4-0 win for Spurs. Hence, despite being unaware of how reality had unfolded, the respondent's strategies proved to be an exact replication of Jose Mourinho's reality during that match.

If we look deeper into the matter, most attacks and shots illustrated in figures 19 and 20 are located from the same channels as the respondents built their strategy around - Tottenham's right flank or Burnley's left. In sum, a result strengthening the practical significance of how efficiently one can locate crucial playing patterns with the artifact, as it took the respondents approximately 20-30 minutes to assess and assemble a game-plan similar to what Jose Mourinho did.

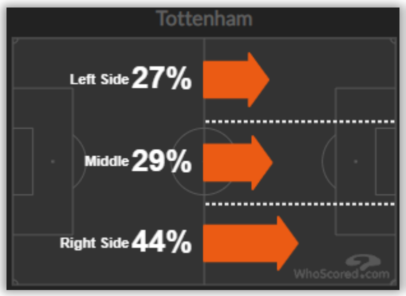


Figure 19: Attack sides.

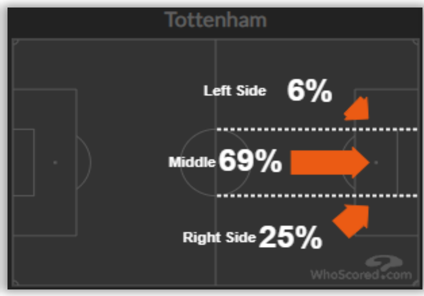


Figure 20: Shot directions.

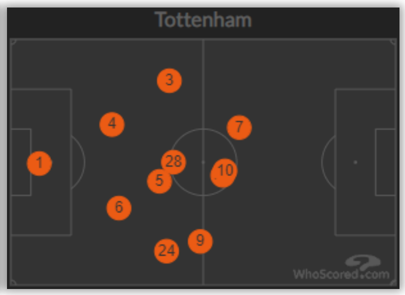


Figure 21: Average positions.

Whereas the players positioned towards the right and the center of the field in figure 21 generated all four assists (A) and goals (G) - (7. *Son*: AA, 10. *Kane*: G, 27. *Moura*: G and 9. *Bale*: GGA). Thus, at last, understat.com (2021) visualized the majority of Tottenham's chances (circles) and goals (stars) to come from the center and slightly towards Burnley's proposed weak side.

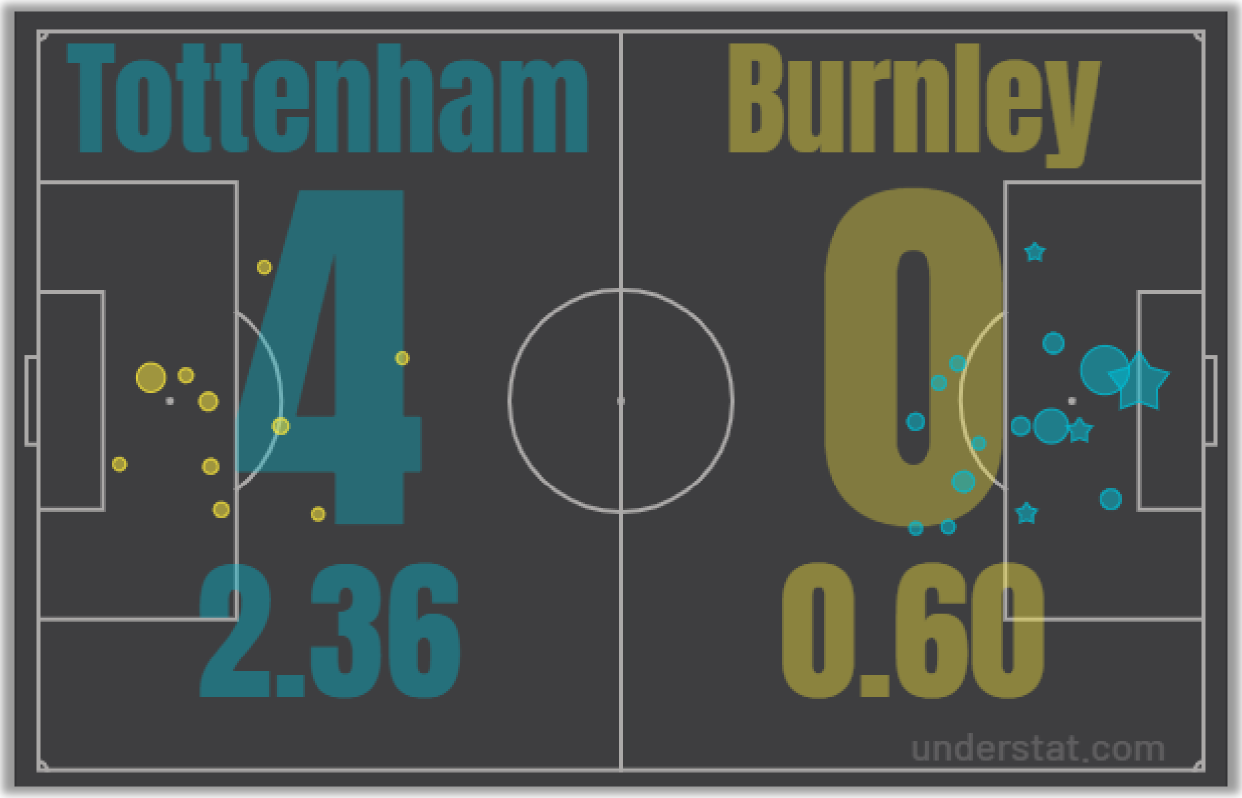


Figure 22: Match report retrieved from understat.com (2021).

Another interesting feature is how the respondents have managed to line up (see Table 12) the same players as Mourinho in almost the same positions, as illustrated in figure 21. This emphasizes how the respondents were able to predict realistic patterns. For example, they prioritized defenders with high ball retention and possession skills, as they expected to recycle the ball to gain an advantage out of the opposition's low press. In accordance, if we compare the player's average positions in figure 21 to the heatmap in figure 18, we can see how this observation played out in reality as most touches are centralized at the same position as their defender's average position. Based on this, the author argues there is most definitely a potential advantage in locating these patterns through the application, especially regarding how it speeds up the process, as expressed by respondent 1:

> Based on all the various data tools we possess today, I assume this would be a much more efficient way of approaching tactics in the future (respondent 1).

In summary, as all respondent explicitly expressed their willingness to apply the web application in their work, the author believes the study to partake in the digital transformation of sustainable big data solutions for soccer tactics that potentially can generate business value in the future.

7. Conclusion and Future Work

The overarching aim of this thesis is to contribute to the socio-technical challenges of how data analytics can simplify the tactical decision-making processes in soccer. As a result, the author has constructed a web application for tactical decision-making by following a design science research (DSR) paradigm. Thus, in this thesis, a data-driven system based on polar charts is introduced as a provider of adjustable pre-constructed templates, processing raw data from StatsBomb to create holistic data visualizations identifying given prospects and tactical patterns according to preferences.

Drawn from the consensus among experts testing the web application, the author concludes the system to have shown great potential in generalizing the strategic process of identifying tactical patterns. Additionally, this results in strengthening the practical significance of how efficient the artifact is to locate a proper strategy, as it took the respondents - unaware of their actual reality - approximately 20-30 minutes to assess and assemble a game-plan almost congruent to the 'the special' Jose Mourinho. Subsequently, as the author believes this domain an uncharted territory, the study contributes to a deeper comprehension of big data's potential impact on soccer tactics.

Despite a promising outcome, as with any contribution of this type, the sole intention of the proposed web application is to serve as a blueprint for future work. As a result, researchers can practice their discretion to vary what is proposed or submit and achieve improvements. In accordance, an implication extending to the proven necessity of contextualizing performance indicators on the end-user premises arises as some metrics tend to contradict personal opinions. To these ends, it would be interesting to look further into how the artifact could adapt to other relevant features in future research. For example, heat maps, pass maps, physical attributes, or chemistry team builders. Potentially, this can generate different hypotheses for later testing, such as testing the assumption that the majority of a team's chance creation aligns with the position of the player with the highest score of successful dribbles. If true, test if the player with the highest successful tackles is the best receipt against such a player, or alternatively, if other attributes are more effective against the opposition's most dangerous weapon.

Finally, it would be interesting to see how the artifact could adapt into reality by observing a soccer club while utilizing the system for further proof-of-use and proof-of-value analyses. Otherwise, as similar big data systems and applications expand, they will inevitably alter longstanding conceptions about decision making, competitive strategy formulation, management practices, and value creation in soccer. Hence, the value that emerges through this thesis proposed ecosystem may pave us to the forefront of an entirely new strand within the analytics industry. One that potentially will attract further attention in the upcoming years.

8. References

Akkermans, H., & van Helden, K. (2002). Vicious and virtuous cycles in ERP implementation: a case study of interrelations between critical success factors. *European Journal of Information Systems, 11*(1), 35-46.

Anderson, C., & Sally, D. (2013). *The Numbers Game: Why Everything You Know About Football is Wrong*: Penguin Books Limited.

Angular. (2021). Retrieved from https://angular.io/

Arnott, D., Lizama, F., & Song, Y. (2017). Patterns of business intelligence systems use in organizations. *Decision Support Systems, 97*, 58-68.

Baskerville, R. L., & Myers, M. D. (2015). Design ethnography in information systems. *Information Systems Journal, 25*(1), 23-46.

Behravan, I., Zahiri, S. H., Razavi, S. M., & Trasarti, R. (2019). Finding Roles of Players in Football Using Automatic Particle Swarm Optimization-Clustering Algorithm. *Big data, 7*(1), 35-56.

Bose, R. (2004). Knowledge management metrics. *Industrial management & data systems*.

Brynjolfsson, E., & McElheran, K. (2019). Data in Action: Data-Driven Decision Making and Predictive Analytics in US Manufacturing. *Available at SSRN 3422397*.

Caban, J. J., & Gotz, D. (2015). Visual analytics in healthcare–opportunities and research challenges. In: Oxford University Press.

Creswell, J. W. (2014). *Research design : qualitative, quantitative, and mixed methods approaches* (4th ed.; International student ed. ed.). Los Angeles, Calif: SAGE.

Davidson, C. (2009). Transcription: Imperatives for qualitative research. *International journal of qualitative methods, 8*(2), 35-52.

de Vaus, D. (2001). The context of design. *Research design in social research, 279*.

Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. (2019). *Actions speak louder than goals: Valuing player actions in soccer*. Paper presented at the Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.

Dick, U., & Brefeld, U. (2019). Learning to rate player positioning in soccer. *Big data, 7*(1), 71-82.

Erraissi, A., & Belangour, A. (2018). Data sources and ingestion big data layers: meta-modeling of key concepts and features. *International Journal of Engineering & Technology, 7*(4), 3607-3612.

FBref.com. (2021). Retrieved from https://fbref.com/en/

Fischer, C., Winter, R., & Wortmann, F. (2010). Design theory. *Business & Information Systems Engineering, 2*(6), 387-390.

Goes, F., Meerhoff, L., Bueno, M., Rodrigues, D., Moura, F., Brink, M., . . . Torres, R. (2020). Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sport Science*, 1-16.

Goldkuhl, G. (2004). Design theories in information systems-a need for multi-grounding. *Journal of Information Technology Theory and Application (JITTA), 6*(2), 7.

Goldkuhl, G. (2012). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems, 21*(2), 135-146.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS quarterly*, 337-355.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*: Sage.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management, 53*(8), 1049-1064.

Hassan, A., Akl, A.-R., Hassan, I., & Sunderland, C. (2020). Predicting Wins, Losses and Attributes' Sensitivities in the Soccer World Cup 2018 Using Neural Network Analysis. *Sensors (Basel, Switzerland), 20*(11), 3213. doi:10.3390/s20113213

Heggernes, T. A. (2017). *Digital forretningsforståelse : fra store data til små biter* (2. utg. ed.). Bergen: Fagbokforl.

Hellevik, O. (2002). Forskningsmetode i Sosiologi og Statsvitenskap. Oslo. In: Universitetsforlaget.

Hennessy, L., & Jeffreys, I. (2018). The current use of GPS, its potential, and limitations in soccer. *Strength & Conditioning Journal, 40*(3), 83-94.

Hevner, A., March, S. T., Park, J., & Ram, S. (2004). Design science research in information systems. *MIS quarterly, 28*(1), 75-105.

Horváth, I. (2007). *Comparison of three methodological approaches of design research.* Paper presented at the DS 42: Proceedings of ICED 2007, the 16th International Conference on Engineering Design, Paris, France, 28.-31.07. 2007.

Hutchby, I., & Wooffitt, R. (1998). Data and Transcription Techniques. *Conversation Analysis: Principles, Practices and Applications*, 73-142.

Instatsport. (2020). Retrieved from https://instatsport.com/

Israel, M., & Hay, I. (2007). Research ethics for social scientists. *Social Work & Social Sciences Review, 12*(3), 79-83.

Ivanov, T., & Singhal, R. (2018). *Abench: Big data architecture stack benchmark.* Paper presented at the Companion of the 2018 ACM/SPEC International Conference on Performance Engineering.

Jacobsen, D. I. (2015). *Hvordan gjennomføre undersøkelser? : innføring i samfunnsvitenskapelig metode* (3. utg. ed.). Oslo: Cappelen Damm akademisk.

Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., . . . Gani, A. (2014). Big data: survey, technologies, opportunities, and challenges. *The scientific world journal, 2014*.

Kharrat, T., McHale, I. G., & Peña, J. L. (2020). Plus–minus player ratings for soccer. *European Journal of Operational Research, 283*(2), 726-736.

Kirk, A. (2016). *Data Visualisation: A Handbook for Data Driven Design*: Sage Publications Ltd.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering–a systematic literature review. *Information and software technology, 51*(1), 7-15.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management, 34*(3), 387-394.

Lewis, M. (2004). *Moneyball: The art of winning an unfair game*: WW Norton & Company.

Llana, S., Madrero, P., Fernández, J., & Barcelona, F. (2020). *The right place at the right time: Advanced off-ball metrics for exploiting an opponent's spatial weaknesses in soccer*. Paper presented at the Proceedings of the 14th MIT Sloan Sports Analytics Conference.

Lucey, P. (Writer) & FourFourTwo (Director). (2017). The Numbers Game | How Data Is Changing Football | Documentary. In *The Numbers Game | How Data Is Changing Football | Documentary*: FourFourTwo.

MacLean, L. M., Meyer, M., & Estable, A. (2004). Improving accuracy of transcripts in qualitative research. *Qualitative health research, 14*(1), 113-123.

Magnello, M. E. (2012). Victorian statistical graphics and the iconography of Florence Nightingale's polar area graph. *BSHM Bulletin: Journal of the British Society for the History of Mathematics, 27*(1), 13-37.

McElheran, K., & Brynjolfsson, E. (2017). The rise of data-driven decision-making is real but uneven. *IEEE engineering management review, 45*(4), 103-105. doi:10.1109/EMR.2017.8233302

McHale, I. G., & Relton, S. D. (2018). Identifying key players in soccer teams using network analysis and pass difficulty. *European Journal of Operational Research, 268*(1), 339-347.

Memmert, D., & Rein, R. (2018). Match analysis, big data and tactics: current trends in elite soccer. *German Journal of Sports Medicine/Deutsche Zeitschrift fur Sportmedizin, 69*(3).

Mikalef, P., Pappas, I., Krogstie, J., & Pavlou, P. A. (2020). *Big data and business analytics: A research agenda for realizing business value*: Elsevier.

Mitchell, R. (2018). *Web scraping with Python: Collecting more data from the modern web*: " O'Reilly Media, Inc.".

Munkvold, B. E. (1999). Challenges of IT implementation for supporting collaboration in distributed organizations. *European Journal of Information Systems, 8*(4), 260-272.

Myers, M. D., & Avison, D. (2002). *Qualitative research in information systems: a reader*: Sage.

Nelli, F. (2015). *Python data analytics: Data analysis and science using PANDAs, Matplotlib and the Python Programming Language*: Apress.

Neuman, W. L. (2014). *Basics of social research*: Pearson/Allyn and Bacon.

Neyer, R. (2017). Sabermetrics. Retrieved from https://www.britannica.com/sports/sabermetrics/The-rise-of-advanced-statistics

O'Connor, S., Waite, M., Duce, D., O'Donnell, A., & Ronquillo, C. (2020). Data visualization in health care: The Florence effect. In: Wiley Online Library.

Oates, B. J. (2005). *Researching information systems and computing*: Sage.

OECD. (2015). *Frascati Manual 2015*.

Oliver, D. G., Serovich, J. M., & Mason, T. L. (2005). Constraints and opportunities with interview transcription: Towards reflection in qualitative research. *Social forces, 84*(2), 1273-1289.

Oussous, A., Benjelloun, F.-Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences, 30*(4), 431-448.

Palanivel, K. (2019). Modern Network Analytics Architecture Stack to Enterprise Networks. *International Journal for Research in Applied Science & Engineering Technology (IJRASET), 7*(4), 263-280.

Pappalardo, L., Cintia, P., Rossi, A., Massucco, E., Ferragina, P., Pedreschi, D., & Giannotti, F. (2019). A public data set of spatio-temporal match events in soccer competitions. *Scientific data, 6*(1), 1-15.

Pappas, I. O., Mikalef, P., Giannakos, M. N., Krogstie, J., & Lekakos, G. (2018). Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies. In: Springer.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems, 24*(3), 45-77.

Peral, J., Maté, A., & Marco, M. (2017). Application of data mining techniques to identify relevant key performance indicators. *Computer Standards & Interfaces, 54*, 76-85.

Peralta Alguacil, F., Fernandez, J., Piñones Arce, P., & David, S. (2020). Seeing in to the future: using self-propelled particle models to aid player decision-making in soccer. *Sports Analytics Conference*, 23.

Playermaker. (2020). Retrieved from https://playermaker.com/

Pollard, R. (2002). Charles Reep (1904-2002): pioneer of notational and performance analysis in football. *Journal of sports sciences, 20*(10), 853-855. doi:10.1080/026404102320675684

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data, 1*(1), 51-59.

Rangaiah, M. (2020). Role of Business Intelligence in the Sports Industry. Retrieved from https://www.analyticssteps.com/blogs/role-business-intelligence-sports-industry

Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *Springerplus, 5*(1), 1-13. doi:10.1186/s40064-016-3108-2

Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter, 19*(4), 1-34.

Sonnenberg, C., & Vom Brocke, J. (2012). *Evaluations in the science of the artificial–reconsidering the build-evaluate pattern in design science research.* Paper presented at the International Conference on Design Science Research in Information Systems.

Spearman, W. (2018). *Beyond expected goals.* Paper presented at the Proceedings of the 12th MIT sloan sports analytics conference.

StatsBomb. (2020). Retrieved from https://statsbomb.com/teams/

Sykes, J., Paine, N. (2016). How One Man's Bad Math Helped Ruin Decades Of English Soccer. Retrieved from https://fivethirtyeight.com/features/how-one-mans-bad-math-helped-ruin-decades-of-english-soccer/

understat.com. (2021). Tottenham 4 - 0 Burnley. Retrieved from https://understat.com/match/14693

Van Haaren, L. B. J. (2020). Player Chemistry: Striving for a Perfectly Balanced Soccer Team. *Sports Analytics Conference*.

Venable, J. (2006). *The role of theory and theorising in design science research.* Paper presented at the Proceedings of the 1st International Conference on Design Science in Information Systems and Technology (DESRIST 2006).

Veo. (2020). Retrieved from https://www.veo.co/soccer-camera/

Walsham, G. (1995). Interpretive case studies in IS research: nature and method. *European Journal of Information Systems, 4*(2), 74-81.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly,* xiii-xxiii.

whoscored.com. (2021). Positional Report. Retrieved from https://www.whoscored.com/Matches/1485361/Live/England-Premier-League-2020-2021-Tottenham-Burnley

Wyscout. (2020). Retrieved from https://wyscout.com/

Yin, R. K. (2017). *Case study research and applications: Design and methods*: Sage publications.

Appendices

A – Expert Review

Background questions (5 min)

1. How would you specify your current role with regards to analytics?

2. In accordance with that role, what is your specific skillset?

3. Would you say there exists a data-driven culture in your environment?

Current approach to data analytics (10 min)

4. Which analytical tools have you integrated into your work process? And how do you approach them with regards to tactical decisions?
(e.g., software, stats, whiteboard, etc.)

5. On what grounds have your organization chosen those tools over other tools?
(e.g., cost, knowledge, etc.)

6. To what extent does the current state of these tools suit your needs?

7. How do you progress when data contradicts your personal opinions?

Utilization of the artifact (Expert review – 20 min)

Note: a description (Located in appendix) of the proven metrics should be visual at all time during the review

Restrictions (necessary for validation criteria):

- The final line-up will be compared to the actual line-up, hence..,
- You need to play 4-2-3-1
- You have the same 18 players at your disposal as the actual manager had at that time
- Opposition line-up is real
- Defenders should play as defenders (as they are benchmarked against each other – see the last requirement)
- Midfielders & attackers can be placed from midfield and up.
- Your opposition tend to play a defensive style of play
- Player names are hidden to prevent informants from subjective picks
- All stats reflect a given player's performance up to that date (28/02/2021)
- Some charts make less meaning; this is due to the difference in minutes played
- All players are benchmark towards players in the same position (as they should be evaluated towards the environment they compete in)

Use case

Based on the restriction above…

8. How would you line-up your defense? (Elaborate on your reasonings for each position)

9. How would you line-up your midfield? (Elaborate on your reasonings for each position)

10. How would you line-up your attack? (Elaborate on your reasonings for each position)

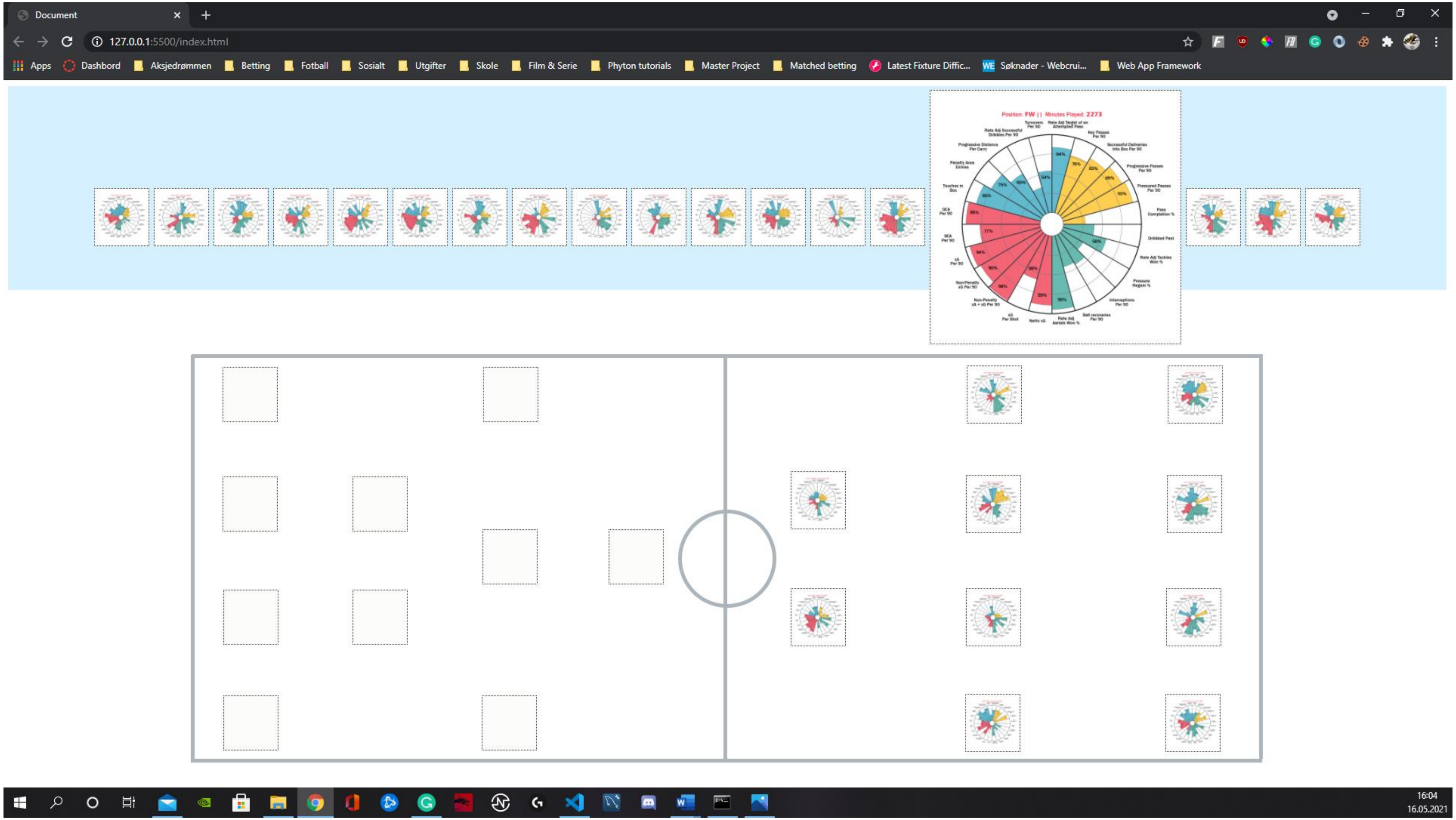
The final output will be embedded here when transcribed

Expert feedback (5 min)

11. Costumed for your needs, would this artifact simplify your current approach to tactics? (Please elaborate on why or why not)

12. Have you seen a similar approach to tactics before?

B – Expert Review Software



C – Artifact Description

This thesis’**s overarching purpose** is to examine the socio-technical challenges of how data analytics is utilized in decision-making processes to decipher the potential of big data in soccer tactics. The main target is to assess/test the utility of the web application illustrated in figure 2 and compare the tool to current work methods. The application is constructed based on the knowledge extracted from an initiating case study (examined a European soccer organization’s approach to data) and a literature review (identifying soccer's most proven metrics). In brief, the tactical tool intends to improve the traditional whiteboard by providing an *all-in-one interface* where player-centric performance data aggregated in a pizza slice chart helps the manager to assess all available players strategically against each other. Ultimately, the complete version of the application merges each player’s position-modified pizza slice chart (Figure 1) within a tactical drag-and-drop whiteboard (Figure 2).

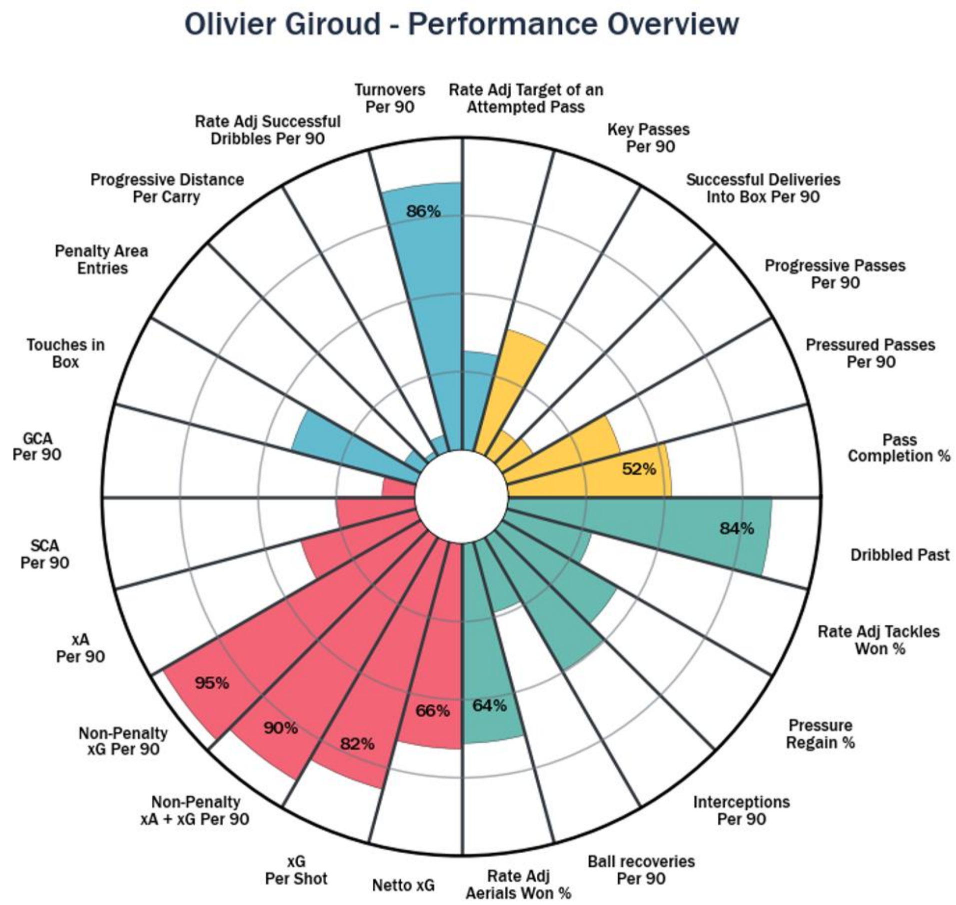
Interpreting the pizza slice chart (figure 1)

In brief, a polar chart is made up of a number of wedges (slices), each representing a proven metric. Each wedge's length corresponds to the selected player’s percentile rank for that metric compared to the players in the same league and position. The percentile rank is the percentage of scores within a dataset equal to or lower than the score. This is reversed for metrics where a lower value is better, such as ‘turnovers.’ In sum, a more extensive bar is always better. Hence, when interpreting Oliver Giroud’s performance overview (Figure 1) so far in the 20/21 season, one immediately notices the amount of filled attacking percentile wedges, indicating Giroud’s outstanding ability to create expected goals per 90 minutes (xG). Another striking attribute describing Giroud’s playstyle is the few touches he generates per 90 minutes in the opposition’s box. Intuitively, drawn from the knowledge above and the looks of Giroud, one would most definitely conclude that this player, looking more like a mountain than an average soccer player, would be scoring the majority of his goal with a header (= few touch) from within the opposition box. In turn, as a manager, you would probably shape your tactics accordingly – blasting in high crosses aiming at his head. Nevertheless, as a contradiction to this highly intuitive opinion based on Giroud’s physics, his aerial percentile wedge only ranks him above the average striker in the air – winning merely 64 % of the aerials, he contends. Being a manager this also tells that you probably would vast your most dangerous weapon by playing a tactic solely based on high crosses, as your striker’s strength probably lies in his position-ability as much as along the ground or in the air.

Each proven metric is defined and color-ordered according to which game phase literature places them.

Proven Metrics

Possession For passing & ball progression	Transition For dribbling & positioning	Attacking For goal-scoring & chance-creation	Defending For defending
Pass Comp % The percentage of attempted passes that successfully reached a teammate	**Turnovers Per 90** Includes every time a player dissposses or miscontrol - where the player loses the ball	**Goal Creation-Action (GCA) Per 90** The two offensive actions directly leading to a chance, such as passes, dribbles and drawing fouls	**Rate Adjusted Tackles Won %** The number of successful tackles a player makes per 90 minute, adjusted for their success rate.
Progressive Passes Per 90 Completed passes that move the ball towards the opponents goal at least 10 yards - excludes passes from the defending 40 % of the pitch	**Succesful Dribbles %** Percentage of a successful attempt at taking on a player and pass them while retaining possesion	**Shot Creation-Action (SCA) Per 90** The two offensive actions directly leading to a chance to shot, such as passes, dribbles and drawing fouls	**Interceptions Per 90** The amount of succsessfully interception per 90 minutes
Passes into 1/3 Per 90 Completed passes that enter the enter the final third of the pitch closest to the goal - not included set pieces	**Rate Adjusted Dribbles Per 90** The number of successful dribbles a player makes per 90 minutes adjusted for their success rate	**Non-Penalty xG Per 90** Expected goals per 90 minute Excluding set pieces such as penalty kicks	**Ball Recoveries Per 90** Every time a player regain possession of a loose ball per 90 minutes
Successful Deliveries into Box Per 90 Completed passes that enter the penalty area of the pitch closest to the goal - not included set pieces	**Progressive Distance Per Carry** The number of yards a player carries the ball towards the opposition goal per carry the player complete in the average 90 minutes	**xA (Expected Assists) Per 90** Expected assist Per 90 included set pieces	**Pressure Regain %** Number of times the team win the ball within 5 seconds after the player applied pressure to opposing player who is receiving, carrying or releasing the ball
Pressured Passes Per 90 Passes made while under pressure from opponent	**Touches in 1/3** The amount of times a player touch the ball in the final third towards the oppenent goal	**Non-Penalty xG + xA Per 90** Non-Penalty expected goal + expexted assist per 90 minute	**Dribbled Past** Number of times a player is dribbled past by an opposing player
KeyPasses Per 90 Passes that directly lead to a shot (assisted shot)	**Touches in Box** The amount of times a player touch the ball in the opponents penalty area	**xG Per Shot** Expected goal per shot the player attempt	**Rate Adj Aerials Won %** Percentage of aerial attempts that is succesfully won, adjusted for their success rate.
	Rate Adj Target of an Attempted Pass The number of times a player successfully recived an attempted pass adjusted for the attempted success rate - indicates a player´s positioning ability	**Netto xA** Actual assists minus expected assists - indicates if a player overperform/underperform	**Clearances Per 90** The amount of times a player clear the ball per 90
	Final Third entries The amount of carries that enter the 1/3 of the pitch closest to the opposition goal	**Netto xG** Actual goals minus expected goals - indicates if a player overperform/underperform	**Blocks Per 90** The amount of times a player blocks the ball per 90
	Penalty Box entries The amount of carries that enter the penalty area of the pitch closest to the opposition goal		

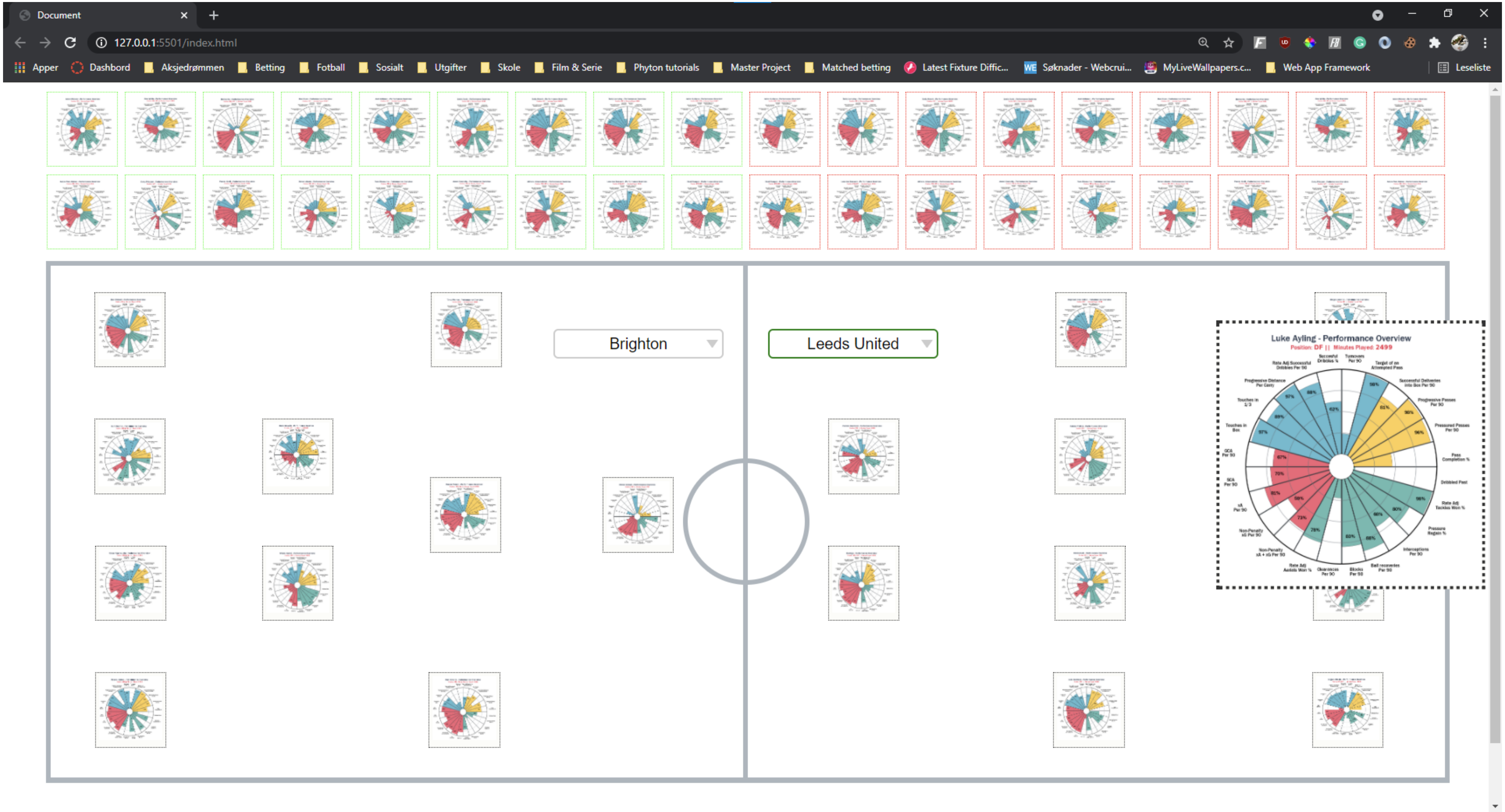


Interpreting the tactical web application (Figure 2)

A proof of the artifact's utility is anchored in its ability to assess Giroud's playstyle and identifying the logic behind his previous performances, as well as how to counter it. For example, by looking into the already stated goal-scorer, one initially tends to acknowledge the low scores he has generated regarding touches in the opposition box, goal creation-actions, and being a target of attempted passes. Taking these remarks into perspective and assess them towards Giroud's generally low passing attributes, few touches/involvements before entering the opposition box, and his low shot creation-action, it is evident that Chelsea does not very much include the striker in build-up plays - even though he is known as a link-up player. One can assume that he preferably operates as a so-called 'fox in the box' (most goals come from tap-ins). If we process this information further and add Giroud's average pressure stats, one should consider countering his abilities with defenders scoring high on stats reflecting good capabilities if pressured. Such as pressured passes, pass completion, turnovers, and a generally high combination of defensive stats to handles Giroud's superior off-ball movement. As illustrated below, a reasonable nemesis for Giroud would be Luke Ayling (current center-back for Leeds).

Another example of the application's utility is identifying the high possibility for Leeds to attack through their left flank. As a result, these players tend to have a higher attacking contribution (reflected in their red bars).

In sum, the artifact aims at putting the user in a better position than ever to outmaneuver the opposition tactically.


Document
127.0.0.1:5501/index.html
Apper
Dashbord
Aksjedrømmen
Betting
Fotball
Sosialt
Utgifter
Skole
Film & Serie
Phyton tutorials
Master Project
Matched betting
Latest Fixture Diffic...
Søknader - Webcrui...
MyLiveWallpapers.c...
Web App Framework
Leseliste
Brighton
Leeds United
Luke Ayling - Performance Overview
Position: DF || Minutes Played: 2499