The Potential of Big Data Analytics for Decision Support in Sports– The Case of Soccer

Completed Research

Katja Bley
Norwegian University of Science and Technology (NTNU) & TU Dresden
katja.bley@ntnu.no

Martin Holthe Rønningen
University of Agder, Kristiansand
martinholthe@hotmail.no

Paolo Spagnoletti
University of Agder, Kristiansand
paolo.spagnoletti@uia.no

Ilias O. Pappas
University of Agder, Kristiansand & Norwegian University of Science and Technology (NTNU)
ilias.pappas@uia.no

Abstract

Compared to other popular sports, soccer is quite a fluid sport with high degrees of freedom, which is why evaluating games or single players' performance is more demanding and complex. Even though the availability of different types of soccer data has increased steadily over the last few years, the process of strategic and tactical decision-making is still traditionally done by single specialists. In this context, the potentials of big data analytics (BDA) provide an advantage in automatically assessing and assembling data from various sources and generating insights that exceed the analytical capability of individual human experts. We focus on this potential by developing a design science research-based BDA artifact, which aims at providing strategic decisional support for managers and experts. In several evaluations with different soccer and sports experts, we are able to show advanced usability of the artifact's instantiation over traditional tools provided by large stats providers.

Keywords

Big data analytics, decision support, strategic support, soccer.

Introduction

Sports are driven by an ongoing competition, and especially in group sports strategic superiority over the opponent can lead to a victory. Soccer, for instance, is a more fluid sport than other popular sports (e.g., baseball) and, with twenty-two players always committed, it has more moving elements than basketball or ice hockey. Thus, from a mathematical perspective, soccer has more degrees of freedom than other sports, making it difficult to evaluate the game or single players' performance using a small number of metrics. Research has focused more on specific in-game data and how to approach them, and less on how soccer clubs can assemble and utilize them through big data analytics (BDA) (Goes et al. 2021). 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 valorise these data.

For the analysis, tactical identifiers (metrics) exist, such as press height, passes per defensive action, action zones, usage rates for individual players, or pitch areas (StatsBomb 2020). Nevertheless, as management is uncertain of these stats and most of the tactical work is traditionally done by the coaching staff – limiting both their team analysts and external researchers from exploring new tactical disruptions. Further, Memmert and Rein (2018) justify BDA's 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 in collaboration to support the team's decision-making processes.

Thus, a shift from traditional, more qualitative analysis methods to modern data-driven game analysis techniques is needed, as data-driven decisions are already changing soccer dynamics, from pre-shaping tactics to pinpointing transfer targets (Anderson and Sally 2013). As Goes et al. (2021) claim, modern match analysts require knowledge across the *computer science* and *sports science* domains. To address this research gap and to answer the call for research by Goes et al. (2021), we focus on the socio-technical difficulties on how player-centric metrics can be utilized to simplify and improve decision-making in soccer. By following a design science research (DSR) paradigm proposed by Sonnenberg and vom Brocke (2012), we develop a web application and thereby answer the following research question:

"How can big data analytics be utilized for supporting decision-making processes in soccer?"

To assemble and test the application, a two-phase study is conducted. In the first phase, a conceptual framework was developed, based on the work of Rein and Memmert (2016), which served to define the analytical and technological requirements for the construction of the web application prototype. In the second phase, we instantiated the prototype artifact and applied existing soccer data as an initial evaluation of its applicability. The outcomes are presented with polar charts, which represent personalized data visualizations identifying given prospects and tactical patterns according to the users' preferences. Then, we performed a use case with experts replicating a soccer match to demonstrate the artifact's utility. Our results show that it took the respondents approximately 20-30 minutes to assess and assemble a game-plan for a soccer match. Thus, the study contributes to a deeper understanding and eased application of BDA solutions in soccer, which can be understood as representative for sports that are characterized by high degrees of freedom.

The remainder of the paper is structured as follows. First, we present the theoretical background on different types of data analyses well as the results of the literature review. Afterwards we present the conceptual approach and the underlying design science build-evaluating cycle. We further present the prototypes architectural characteristics in section four which is followed by the artifact's evaluation steps. We end the paper with a discussion of the results and a summary.

Data Analytics in Soccer – State of the Art

To obtain digitalized statistics derived from technical advancements, soccer organizations gather and re-engineer data through self-developed mechanisms or acquire software through well-established online stats providers like Opta, Wyscout, Second Spectrum, STATS, SciSports, and StatsBomb. Such data can be categorised based on their source into *event and tracking data* as well as *coaching and scouting data* (Memmert and Rein 2018; Rein and Memmert 2016). E*vent data* annotates the times and locations of specific events (e.g., passes, shots, and cards) that occur in a game, whereas *tracking data* records the players' locations and the ball at a high frequency using optical tracking systems during games. Such data enable an analyst to gather, capture and contextualize all the various events in soccer from a *coaching* and *scouting* perspective. However, due to the high expenses needed for optical tracking systems, tracking data is mainly available to wealthy leagues or clubs, raising the need for solutions that can be by smaller clubs. Furthermore, a rising issue in this context refers to data privacy as commercial institutions, private clubs, and public research institutions oversee the accumulated data logs. This is rather crucial as professional soccer teams are reluctant to share data due to a possible loss of competitive advantage.

Using the conceptual framework of Rein and Memmert (2016) and review of extant literature we draw conclusions about the most applicable metrics for profiling a player according to tactical decisions. These metrics include chance creation, physics, critical success factors, and player ratings. A strong potential indicator besides the on-ball metrics is the off-ball metrics or off-ball scoring opportunities (Llana et al. 2020; Spearman 2018). It refers to those situations when a player is in a favorable disposition to receive a potential pass. In case of receiving it, the player would likely improve the possession's value. However, due to the limited nature of event data, these metrics only measure a small portion of what actually happens during a soccer game. For instance, during an ordinary match, Barcelona (now Atletico Madrid) striker Luis Suarez possessed 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 actual event data only but should be evaluated by considering the off-ball metrics (Peralta Alguacil et al. 2020). Another concept is the one of *expected variables*, as it presents an enhanced methodology based on the conventional plus-minus differential method to calculate expected events like expected goals (xG), expected points and plus-minus rating. vs. actual events. Kharrat et al. (2020) argue that xG can be more informative than actual goals when evaluating how well a team has played. However, since goals are rare, they do not always reflect a team's performance on the pitch.

Scouts and performance analysts are continuously looking for data-driven tools and metrics to improve the retrieval of talented players with in-demand characteristics. In general, such metrics are measures of quantitative assessments commonly used for reviewing, benchmarking, and tracking performance (Bose 2004). In this context, Dick and Brefeld (2019) proposed an in-depth approach to learn valuations of multiplayer positionings using positional data. For instance, correlations between a dangerousness metric and the traditional performance indicators like 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. These insights could then be integrated into the individual game plan and support a manager in his decisions. Further, Bransen and Van Haaren (2020) argue for an observation of the chemistry between players who have played together in the past. This setting is relevant for a manager who needs to decide on the best possible line-up for an upcoming match. Based on such data, it may be possible to forecast the chemistry between players who have never played together before. This is a particularly relevant context for scouts assessing the fit of a future signing.

Based on the retrieved insights from literature, we extended an existing framework from Rein & Memmert (2016) and summarized 14 concepts of event and tracking as well as coach and scouting data as being relevant for the tactical analysis of soccer data, which served as an initial overview of the research field (Table 1).

Event and Tracking Data				**Coach- and Scouting Data**			
Chance Creation	Physics	Crit. Success Factors	Player Ratings	Offensive Attributes	Defensive Attributes	Roles/ Key Players	Player Recruitment
On-ball/ Technical	Off-ball/ Spatial	Expected Variables		Heat Maps & Pass Maps	Chemistry	Tactical Behavior	

Table 1. Concepts Retrieved from Literature

Further, we conducted an exploratory case study and presented the identified concepts to two experts, who are employees of one of the largest European soccer clubs and involved in strategic decisions, in order to evaluate our results. Based on their feedback on the 14 concepts as being relevant for strategic decision-making and the current possibilities of tactical analysis systems, we identified 24 key performance indicators (KPI) as being decisive for deriving strategic and tactical decisions in soccer. They are divided into four dimensions, related to a game's activity measurement: possession, transition, attacking, and defending. Within these dimensions the 24 KPIs represent distinguished individual measurement scores, which present the resulting underlying framework, and serve as the foundational data poof for the intended artifact. The access is provided via the data base StatsBomb, which is one of the largest and best provider of rich, diversified event data on the market (StatsBomb 2022).

Conceptual Approach

The overall goal of this approach is to develop a web application, that can provide strategic and managerial decision support for soccer tactics based on a provided data pool. Even though online stats providers offer publicly accessible soccer data as well as various tools for the analysis, the potential of such applications as easily accessible tools for an algorithm-based analysis and interpretation of such big data bases remains limited. We focus on this research gap and develop a decision support tool, following a DSR approach. Since this artifact is highly dependent on theoretical as well as conceptual assumptions and relies on several iteration cycles, we follow the build-evaluate research approach proposed by Sonnenberg and vom Brocke (2012) for the rigorous development of a design science prototype artifact (see Figure 1).

Following Goes et al.'s (2021) call for research that enables match analysts to exceed their knowledge in computer science and sport science, we aim at providing a web application for tactical decision-making in soccer (Identify Problem). Thus, we did a literature review and identified the mentioned research problem, Further, we obtained an overview and comprehended which factors represent the frontiers of BDA initiatives in soccer (Eval 1). Based on these results we developed a conceptual overview of relevant factors and identified an initial set of 14 concepts (Table 1) for the assessment and evaluation of players (Design). We further evaluated our results by conducting an exploratory case study with two experts from one of the largest European soccer clubs (Eval 2). Based on their feedback we were able to derive a framework with KPIs that will inform the prototype artifact. Hence, the theoretical foundation of this paper represents the descriptive theory for the design science artifact and refers to the ex-ante evaluation cycle of the prototype.

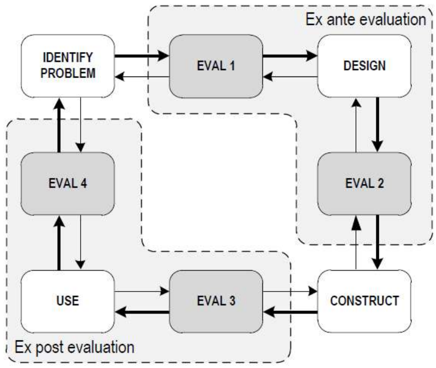

Figure 1. Build-evaluate Cycle by Sonnenberg and vom Brocke (2012)

Based on findings from phase one and the ex-ante evaluation, we developed a web application for tactical analysis, using concrete KPIs that can assess and evaluate the identified challenges and increase the prerequisites for a successful game plan (Construct). This application is created by applying different tools and methods necessary for the technical instantiation of the prototype as web application. For the evaluation of the constructed protype artifact, we ran a user scenario with the data of a soccer player named Patrick Bamford and analyzed his actual performance in season 20/21 with the results of our prototype to see if the retrieved results and charts are meaningful and how a manager could approach a tactical decision (Eval 3). The actual use of the artifact was tested and evaluated by four experts, who were asked to first explain their current approach to tactical decision making and afterwards compare it with the results of our artifact (Use and Eval 4). Based on our derived results, the extensive evaluations and iteration cycles we were able to develop a BDA tool for tactical and managerial decision support in soccer, that is so far missing in the field of IS research.

Construction of the Prototype Artifact

Distributed BDA applications require a complete end to end architecture stack comprising several big data technologies (Ivanov and Singhal 2018). It consists of four conceptual layers.

Ingestion layer: The purpose of the ingestion layer is to allocate and integrate chosen data 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 and Belangour 2018). Further, to route data into a storage solution it must be reproduced from an external source. As ingesting data source, we chose 'StatsBomb' as it represents the most proven and trusted source to retrieve layered soccer data within the analytics community: *"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). The web application extracts, transforms and loads StatsBomb metrics into tabular data tables stored in a MySQL database via a self-made web scraper method using Python (Mitchell 2018).

Data layer: 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 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).

Processing layer: 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 processing. 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 visual analytics layer (Palanivel 2019). Familiar tools and technologies used in the processing layer include PostgreSQL, Apache Spark, Redshift by Amazon etc.

Analytical layer: Finally, the analytical layer is the top layer in the BDA stack, which represents the interface on which the end users interact with the system. Furthermore, the layer involves visualizations such as status reports, dashboards, and business intelligence systems. Hence, the analytics layers' most crucial component is to create a visual representation of the data analysis process, which is easily understandable and manageable by its users.

For the visual representation of the results of soccer analytics, the polar chart was chosen to be the most efficient way to benchmark and visualize a soccer player's ability. It 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 sectors 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.

Use and Evaluation of the Artifact

To test the prototype's functionality, we conducted a proof-of-concept user scenario and provided the web-application artifact with the StatsBomb data of a soccer player named Patrick Bamford, who is playing for Leeds United. We chose his data as he is known for being an overall versatile soccer player, and thereby offers the possibility to test the holistic visual appearance of the artifact. We analyzed his actual performance in season 20/21 with the results provided by the web artifact. To do so we compared the respective results and charts to explore if they are meaningful and how a manager could approach a tactical decision based on them (Eval 3).

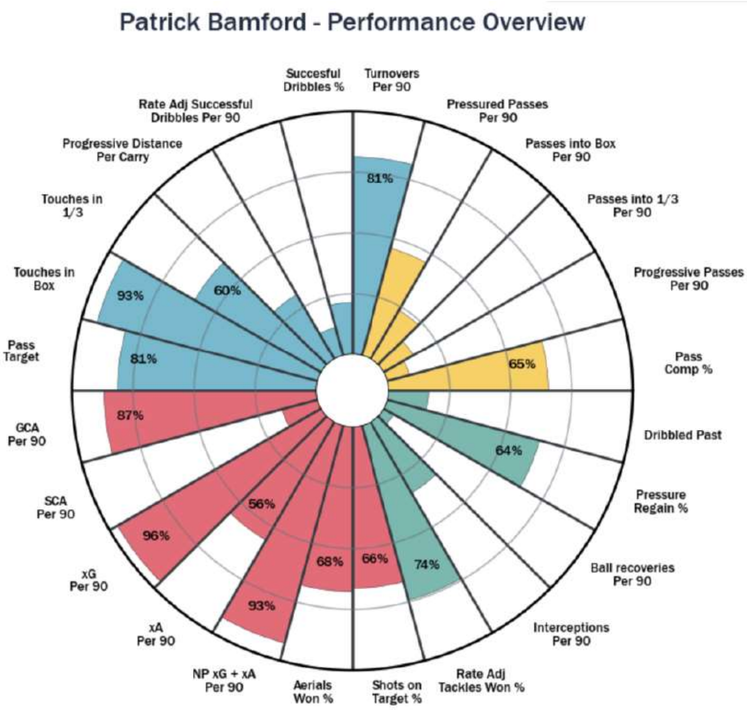

Figure 2. Visual Outcome of the Prototype's Application (Evaluation 3)

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. In sum, a more extensive bar is always better. Hence, when interpreting the players' performance overview according to Figure 2 in the 20/21 season, one immediately notices his outstanding ability to create expected goals per 90 minutes (xG). Another striking attribute describing the player's skillset is the number of touches he generates per 90 minutes in the opposition's box (touches in box). Since Bamford is a relatively tall player (1.85 m), one could conclude that his advantage is scoring his goal with a header from within the opposition box. In turn, a manager would probably shape the tactics accordingly blasting in high crosses aiming at his head. Nevertheless, as a contradiction to this highly intuitive opinion based on the player's physics, his aerial percentile wedge just ranks him above the average striker in the air winning merely 68 of the aerials he contends. This example shows that a tactic decision exclusively based on high crosses might lead to a loss of potential as the player's strength seems to lie along the ground.

The instantiation of the prototype artifact thereby reveals one of its main benefits: due to the direct comparison of the four dimensions with in total 24 performance parameters, it is possible for the user to gain an easily understandable and interpretable overview of a player's strength and weaknesses (Figure 2).

As we were able to show the artifact's general applicability by instantiating it for a single player, we further evaluated the artifact by instantiating it in a complex game scenario (Use), resulting in one pie chart for each of the potentially available players per team (Figure 3 upper part) which can then be placed in a virtual game-plan (Figure 3 lower part). Further, we conducted four in-depth semi-structured interviews with soccer analysts with different professions (Data Consultant, Academy Coach, Sport Scientist, Performance Analyst) of the same large European soccer club as in evaluation step 2 (Eval 4). By receiving feedback of experts with different strategic positions, we aimed at a higher granularity of the artifact's evaluation results. The interviews were conducted in Spring 2021 and lasted between 1 and 2 hours.

Figure 3. Game Scenario for Experts' Evaluation

The interviewees were were asked to apply their regular strategic approach when choosing the desired line-up for a fictional match based on the visual results provided by the artifact (Figure 3), pretending to be the active manager of one of the participating teams. In order to exclude bias, the respondents conducted the assessment without knowing which team they managed (Leeds) or which team they faced (Brighton). Thus, they were able to objectively select which players to include in their strategy.

The interview guide was divided into several phases. Phase 1 serves as an introduction where the study's focus was explained. 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 additional questions. Finally, in phase 4 the experts were given the opportunity to reflect on their assessment and opened for the respondents to correct any misconceptions and provide the necessary additional information.

Experts' Traditional Approach to Data Analytics

Despite all theoretical initiatives developed to standardize analytical practices in soccer, the expert reviews expressed computer- and sport sciences' current approaches causing more inefficiency across a soccer organizations than necessary. As the organizations aim to get a competitive advantage by technological or interpretative advancements, these actions are the main driver for a rising complexity in the quantity and quality of accessible tools from different providers[1].

> "*There is too much work scanning through all these tools we have available concerning the time, and we are in a position where we need to aggregate the ones that will be proven over time.*" (Respondent 2)

Further, as the respondents elaborated on the objectiveness of data, three of the respondents agreed on what happens when data contradicts 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 be 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.*" (Respondent 1)

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 is the limitation of the human eye. In contrast, machines are used to parse and locate key factors in a matter of seconds. In addition, respondent 1 reflects on how he/she 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 individual metrics to identify patterns and exploit this knowledge to predict their collaborative tasks and counter it." (*Respondent 1)

Summarizing, the respondents agreed on the difficulty in choosing an adequate tool for analysis in time, which is on the one hand able to take all required and relevant data into account that is necessary for strategic and tactical decision-making; and which is on the other hand able to visualize and present the results in a meaningful way.

Experts' Approach to the Prototype Artifact

As the respondents followed the abovementioned use case, they indicated no major problems integrating an individualized strategy when choosing a line-up (Figure 3). However, the respondents tend to utilize two distinctive tactical approaches. First, the most tactical qualified respondent located weaknesses within the opponent team, 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 the four defensive players compared to each other, per strengths, weaknesses, threats, and opportunities. The exact process goes for the midfielder and defender. Overall, I think there is very important 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)

[1] For instance: www.wyscout.com, www.instatsport.com, www.playermaker.com, www.veo.co/soccer-camera/

Second, the remaining 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 position in order to increase pressure on their weak 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 before the test, one of the respondents elaborated on how he utilized the polar chart to identify what he deems necessary according to his philosophy and how much trust he could lay in each player:

> *"For the center back position, I value defenders with 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 embellished on how he compared the 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 strong 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)

In order to validate the artifact further, all the respondent's final line-ups were compared to the actual game's line-up from Jose Mourinho. It was remarkable that all the respondents, being very possession-oriented and creative in their philosophy, in contrast to the expert Jose Mourinho, who is known as a defensive strategist, ended up with either a similar or almost the same line-up in less than 30 minutes.

Discussion and Conclusion

The overarching aim of our research approach is to contribute to the socio-technical challenges of how BDA can simplify and support tactical decision-making processes in soccer. Thus, we developed a web application for tactical decision-making by following a build-evaluate design science research cycle by Sonnenberg and vom Brocke (2012). As a result of the artifact's instantiation, polar chart visualizations of singe players are derived which are based on a big data-driven analytical system and which are able to visualize given prospects and tactical patterns according to the prototype users' preferences. Each polar chart consists of 24 performance parameters, combining the advantages of radar charts and pie charts by offering a generic and easily understandable overview of a player's possession, transition, attacking and defending behaviour. As one of the interviewees mentioned, the representation of a polar chart provides an improvement compared to existent tools offered by StatsBomb. Currently, one of the dominant approaches is to analyze data with radar charts, displaying the results as connected data points. Nevertheless, the information content is limited in such visual figures, which is why professional clubs prefer to use their own models for interpretation and analysis. However, by providing polar charts and by illustrating percentiles in bars instead of connected data points, which are not needed or not logically connected, our artifact provides a higher generality for application as well as an improved approach for benchmarking a player's ability in less time.

The statistical revolution of data-driven decision-making has established itself as the nexus navigating modern soccer towards the future. However, as stated by Patrick Lucey, the director of data science at STATS, it is not about gathering the data, as it already exists - it is about creating a universal language of that data (Lucey 2017). Our results thereby amplify Kharrat et al.'s (2020) considerations towards traditional performance indicators simplicity - as these metrics underlying algorithms lack context and a deeper understanding of the situations in which actions are happening. While this is the case, when symbiotic information is provided separately from the different analytical advancements, the interviewed experts found the systematic visualization of the polar charts to present a deeper understanding of how the color-coded aggregation of position-based stats revealed a player's critical playstyle pattern. For example, as the performance analyst applied his regular approach to strategically assess the opposition, he emphasized the advantage of how efficiently he found himself to identified which channels to attack in their defensive line – compared to prior experiences. At the same time, compared to the combination of reviewing and aggregating positional player-centric stats (data point suitable for a player's default position) in a system with similar features, the artifact's application reduced the time-consuming process of scanning and combining software. It amplified the holistic understanding of how coaches efficiently could analyse players' strengths and weaknesses to identify patterns and how the team should exploit this knowledge to their advantage, based on the analysis of large data sets. Additionally, the results strengthen the practical significance of how efficiently the artifact is able to reveal a proper strategy, as it took the respondents approximately 20-30 minutes to assess and assemble a game-plan almost similar to one of the famous experts Jose Mourinho. At the same time, the combination of reviewing positional player-centric stats in a proven system like StatBombs reduced the time-consuming process of scanning the market for appropriate software. Thus, the study contributes to a deeper comprehension of BDA's potential impact on soccer tactics, thus answering the call for more research in this field (Goes et al. 2021). Thereby, our artifact reveals great potential in generalizing the strategic process of identifying tactical patterns in soccer by applying BDA. By developing and presenting the web application artifact, the research question of *"How can big data analytics be utilized for supporting decision-making processes in soccer?"* is addressed and the evaluation revealed a first success of the approach by focusing on the socio-technical difficulties on how player-centric metrics can be utilized to simplify and improve tactical decision-making in soccer.

Besides the practical contribution of the artifact itself, the theoretical contribution of our research is twofold. First, we consider the development process of a strategic and tactical decision-making tool for soccer from a design science research perspective. Thus, based on an ex-ante as well as ex-post evaluation cycle, the development of the web-application artifact is conceptually and theoretically grounded and provides applicants as well as researcher with a possibility to comprehend its theoretical and operating principles. As current strategic decisions in sports are often driven by expert knowledge, we included several soccer experts from various domains in different development phases to ensure high-level feedback for the web application artifact. Second, drawn from the consensus among experts testing the web application, we provide insights into the theoretical application of BDA in soccer for the derivation of strategic decision. Thus, our approach is able to demonstrate the potential of applying BDA in team sports, thereby offering new opportunities for strategic decision-making as well as future research.

As a practical implication of our artifact, all respondents emphasized the importance of using universal syntax and semantics, however agreed on how the artifact tends to be capable of telling the same story to both, a coach and an analyst: *"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)*. Thus, our artifact can be regarded as a first attempt in bridging the gap between the complexity that lies in BDA methods and the requirement of providing an easily accessible tool for strategic analysis in sports. To this end, our research approach provides an initial attempt for big data driven decisional guidance in soccer and our design science-based prototype can be regarded as assisting managers and analysts to recognize a player's performance in relation to potential opponents. Thus, a more streamlined process of consuming tactical knowledge is feasible and the possibility of deriving tactical and strategic decisions is possible.

Like every research project, our approach has limitations. First, the small number of qualitative interviews limits the representativity of our results. Future research should therefore focus on additional evaluation studies and more interviews with different experts in this field. Further, the provided data set of the soccer match could be a limitation as we did not test the artifact's usability with alternative match data yet. However, this is subject of our future research in this area and should also be considered as an opportunity to reveal the artifacts further potential.

REFERENCES

Anderson, C., and Sally, D. 2013. *The Numbers Game: Why Everything You Know about Soccer Is Wrong*, New York: Penguin Books.

Bose, R. 2004. "Knowledge Management Metrics," *Industrial Management & Data Systems* (104:6), pp. 457–468. (https://doi.org/10.1108/02635570410543771).

Bransen, L., and Van Haaren, J. 2020. "Player Chemistry: Striving for a Perfectly Balanced Soccer Team," in *Proceedings of the 14th MIT Sloan Sports Analytics Conference*.

Dick, U., and Brefeld, U. 2019. "Learning to Rate Player Positioning in Soccer," *Big Data* (7:1), pp. 71–82. (https://doi.org/10.1089/big.2018.0054).

Erraissi, A., and Belangour, A. 2018. "Data Sources and Ingestion Big Data Layers: Meta-Modeling of Key Concepts and Features," *International Journal of Engineering & Technology* (7:4), pp. 3607–3612. (https://doi.org/10.14419/ijet.v7i4.21742).

Goes, F. R., Meerhoff, L. A., Bueno, M. J. O., Rodrigues, D. M., Moura, F. A., Brink, M. S., Elferink-Gemser, M. T., Knobbe, A. J., Cunha, S. A., Torres, R. S., and Lemmink, K. A. P. M. 2021. "Unlocking the Potential of Big Data to Support Tactical Performance Analysis in Professional Soccer: A Systematic Review," *European Journal of Sport Science* (21:4), Routledge, pp. 481–496. (https://doi.org/10.1080/17461391.2020.1747552).

Ivanov, T., and Singhal, R. 2018. "ABench: Big Data Architecture Stack Benchmark," in *Companion of the 2018 ACM/SPEC International Conference on Performance Engineering*, Berlin Germany: ACM, April 2, pp. 13–16. (https://doi.org/10.1145/3185768.3186300).

Kharrat, T., McHale, I., and Peña, J. L. 2020. "Plus-Minus Player Ratings for Soccer," *Eur. J. Oper. Res.* (https://doi.org/10.1016/j.ejor.2019.11.026).

Llana, S., Madrero, P., and Barcelona, J. F. F. 2020. "The Right Place at the Right Time: Advanced off-Ball Metrics for Exploiting an Opponent's Spatial Weaknesses in Soccer," in *Proceedings of the 14th MIT Sloan Sports Analytics Conference*. (https://www.semanticscholar.org/paper/The-right-place-at-the-right-time%3A-Advanced-metrics-Llana-Madrero/8fc0d83684301b6ba2aa459bbed548515e2462f2).

Memmert, D., and Rein, R. 2018. "Match Analysis, Big Data and Tactics: Current Trends in Elite Soccer," *German Journal of Sports Medicine* (2018:03), pp. 65–72. (https://doi.org/10.5960/dzsm.2018.322).

Mitchell, R. E. 2018. *Web Scraping with Python: Collecting More Data from the Modern Web*, (Second edition.), Sebastopol, CA: O'Reilly Media.

Palanivel, K. 2019. "Modern Network Analytics Architecture Stack to Enterprise Networks," *International Journal for Research in Applied Science and Engineering Technology* (7:4), p. 2634.

Peralta Alguacil, F., Pinones Arce, P., Sumpter, D., and Fernandez, J. 2020. "Seeing in to the Future: Using Self-Propelled Particle Models to Aid Player Decision-Making in Soccer," in *Proceedings of the 14th MIT Sloan Sports Analytics Conference*.

Rein, R., and Memmert, D. 2016. "Big Data and Tactical Analysis in Elite Soccer: Future Challenges and Opportunities for Sports Science," *SpringerPlus* (5:1), p. 1410. (https://doi.org/10.1186/s40064-016-3108-2).

Sonnenberg, C., and vom Brocke, J. 2012. "Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research," in *Design Science Research in Information Systems. Advances in Theory and Practice* (Vol. 7286), K. Peffers, M. Rothenberger, and B. Kuechler (eds.), Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 381–397. (https://doi.org/10.1007/978-3-642-29863-9_28).

Spearman, W. 2018. "Beyond Expected Goals," in *Proceedings of the 12th MIT Sloan Sports Analytics Conference*. (https://www.bibsonomy.org/bibtex/bfe70febc030c299a9d1b9ffa0b95fb5).

StatsBomb. 2022. "StatsBomb | Football Like Never Before," *StatsBomb*. (https://statsbomb.com/, accessed February 13, 2022).