Development of a Strategic Decisional Support System for Soccer

Katja Bley[1,2][0000-0001-8365-8905], Martin Holthe Rønningen[3], Raoul Hentschel[2], Paolo Spagnoletti[3] and Ilias Pappas[1,3]

[1] Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway
[2] Business Information Systems, esp. IS in Trade and Industry, TU Dresden, Dresden, Germany
[3] Department of Information Systems, University of Agder, Kristiansand, Norway
katja.bley@ntnu.no
martinholte@hotmail.no
{paolo.spagnoletti,ilias.pappas}@uia.no

Abstract. Soccer, in comparison to other popular sports, is a fluid game with high degrees of freedom, which makes analyzing games or the performance of individual players more difficult and challenging. In this context, big data analytics offers an advantage in automatically assessing and synthesizing data from multiple sources, as well as delivering insights that go beyond the analytical capabilities of individual human specialists. We focus on this opportunity by creating a design science research approach to a big data analytics web application artifact that intends to provide strategic decisional guidance support for managers and specialists. The prototype provides the visual presentation of single players' strength and weaknesses in polar charts and can therefore be used to support the tactical and strategic decisions about upcoming game set ups. Thus, we are able to show the potential of big data based analytical tools for strategic decisions in sport and offer an alternative to existing, merely intuitive approaches by experts.

Keywords: Big Data Analytics, Decisional Support, Soccer, Sports.

1 Introduction

In light of the ongoing digital transformation, big data is becoming relevant to fields that have not yet been affected by the potentials of this analytical method [1]. One of such fields is the sport of soccer [2]. As a team sport, it is typically characterized by a dynamic gameplay with a high degree of freedom, making it more challenging to evaluate and to analyze the contribution of single players individually. Thus, the potential and performance of players need to be assessed by considering several interdepending factors simultaneously. Currently, existing approaches often rely on experts' knowledge, as those are able to evaluate potential based on their expertise.

The data, which is used for such analytical approaches can be categorized into *tracking and event* data as well as *coaching and scouting data* [3, 4]. Tracking data records the players' locations and the ball at a high frequency using optical tracking systems

during games, whereas event data annotates the times and locations of specific events (e.g., passes, shots, and cards) that occur in a game. Such data enable an expert or analyst to gather, capture and contextualize all the various events in soccer from a *coaching and scouting* perspective. Several providers for soccer data evolved during the last years. Mostly online providers, like Opta, Wyscout, Second Spectrum, STATS, SciSports, or StatsBomb are examples for data sources, where professional clubs and managers can access large amounts of games, players, or stats. Those stats-based analyses are especially relevant to coaches or managers when it comes to potential contractual takeovers of players from other soccer clubs. Even though the availability of different types of soccer data has increased steadily over the last few years, the process of strategic decisions (e.g., line-up, substitute players) is still traditionally done by single specialists, like coaches. In this context, the potentials of big data analytics (BDA) provide an opportunity to automatically assess and assemble data from various sources, thus, generating insights that exceed the analytical capability of individual human experts.

As Goes et al. [2] propose a closer collaboration between the sports and computer sciences to reveal the potential of big data in supporting tactical performance analysis, we follow their call for research and develop a web application prototype, which can be used for strategic decisional support in soccer. As the development of our artifact is characterized by a higher level of uncertainty regarding a useful and appropriate application and usefulness, we decided to include several evaluation cycles. Thus, we build the web application following a design science research (DSR) cycle proposed by Sonnenberg and vom Brocke [5] which is characterized by ex-ante and ex-post construction evaluations. We thereby address the socio-technical challenge of how BDA approaches can be used to support human experts' decisions in complex sport game settings.

The remainder of this paper is structured as follows. First, we describe the conceptual approach of our artifact followed by the design of the prototype artifact. Further, we present its application and evaluation and based on this its theoretical and practical implications. We end with an outlook on future research opportunities.

2 Design of the Artifact

For a rigorous development process of the prototype artifact, we followed the build-evaluate cycle by Sonnenberg and vom Brocke [5] (Fig. 1). Initially, we identified the opportunity of a BDA-based web application for tactical decision-making in soccer (Identify Problem) and were further able to confirm a research gap in this context within relevant literature (Eval 1). Based on the derived results we developed a conceptual overview of relevant factors (Design). We further evaluated our derived results by conducting an exploratory case study with two experts from one of the largest European soccer clubs (Eval 2). Their feedback enabled us to propose a framework with 24 KPIs that will inform the prototype artifact on the data level.

Based on findings from the ex-ante evaluation, we were able to develop the first prototype of the BDA web application artifact (Construct). Having access to one of the largest data bases for soccer stats, we were able to test the applicability of our prototype (Eval 3). Further, we asked four experts with different professions in soccer to apply

the artifact and use it for strategic game plan decisions (Use). Based on their feedback and were able to evaluate the prototype artifact (Eval 4) and to implement their recommendations for the BDA prototype for tactical and strategic decision support in soccer, that is so far missing in the field of IS research.

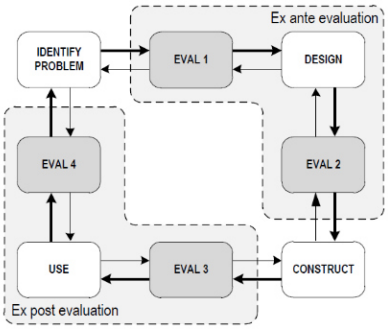

Fig. 1. Build-evaluate cycle by Sonnenberg and vom Brocke (2012)

The BDA prototype was developed by using a complete end to end architecture stack, comprising several big data technologies [6]. It consists of four conceptual layers building on top of each other.

Ingestion layer: The ingestion layer's aim is to allocate and prepare data for the later layers' specialized tools and technologies. [7]. Further, to direct data into a storage solution it must be reproduced from an external source. Thus, we chose 'StatsBomb' as being one of the most proven and trusted source to retrieve layered soccer stats and data within the analytics community [8]. The prototype 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 [9].

Data layer: The data layer represents the backend of the entire system and is generally referred to as the most fundamental layer in the analytical stack. This layer manages the modeling process, which shapes and organizes data to support analytics, in addition to storing all raw data from various data sources provided by the external database [9]. This process grants users to alter data for selective querying [10].

Processing layer: The processing layer begins the actual processing after the pipeline has assigned the data sources and turned them into a desirable ABT stored in the MySQL database. Analysts transform a massive volume of data into meaningful data marts before the final visual analytics layer, making this step probably the most important in the end-to-end big data technology stack. [10]. 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, and it represents the system's interface with end users. Visualizations such as status reports, dashboards, and business intelligence systems are also included in the layer. As a result, the analytics layers' most important function is to generate a visual representation of the data analysis process that users can understand and manage.

When implementing the prototype we built on the representation as a web application, which provides a state-of-the-art user experience, that is closely related to existing approaches, but still offers different functionality. For instance, instead of a radar chart output, which is available at platforms like StatsBomb.com, the polar chart was selected as the most effective technique to benchmark and depict a soccer player's ability in the instance of soccer analytics. It's a combination between a bar chart (using length as a pre-attentive technique) and a pie chart (radial in nature, narrower at the center where segments increase towards the top, i.e., polar coordinates). In fact, the degree of each sector, like in the pie chart, provides percentage data indicated by the categorization in relation to the total. The numerical value of that category is represented by the circular extension in the bar chart.

3 Evaluation and Relevance

The instantiation of the prototype artifact results in a visual representation of a single player's skills in a polar chart, which is based on 24 KPIs about his personal performance. Figure 2 shows a combined output of such polar charts for two soccer teams in a game set up. Based on the visual representation (bar length and percentage), tactical and strategic decisions can be made for potential game plans depending on the individual and situational strengths.

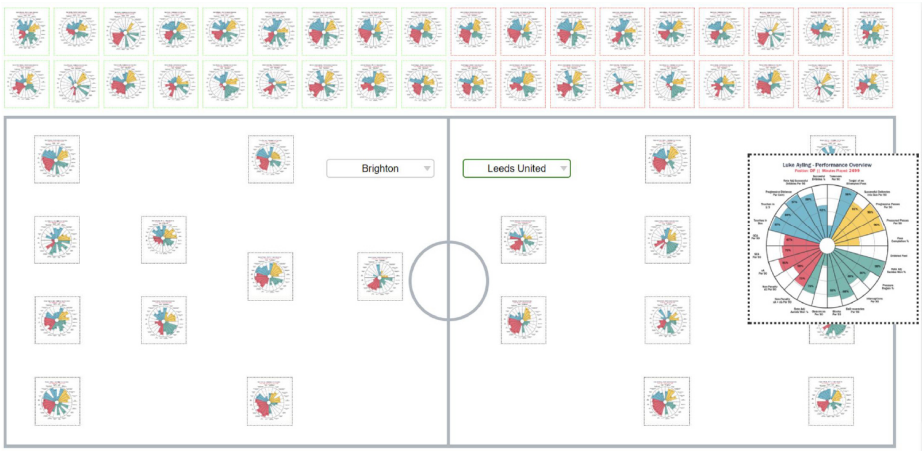

Fig. 2. Graphical output generated by the prototype's web application

As an evaluation of the prototype's usability, we conducted four in-depth semi-structured interviews with soccer analysts (Data Consultant, Academy Coach, Sport Scientist, Performance Analyst). The experts were presented the players' data sets of two anonymized teams, and they were asked to create a game set up while analyzing the individual players' output from our prototype. Afterwards, we evaluated their derived game set ups with the actual game plan from the soccer match of both teams in 2021 (Brighton vs. Leeds United). It was remarkable that all experts - regardless of their

actual profession – were able to develop a game plan, which was almost congruent to one of the leading specialists in this area. As stated by the interviewees, one of the prototype primary benefits is the possibility of a direct comparison of the 24 BDA-based performance parameters for every player. Thereby, the user can easily understand and interpret a player's strengths and limitations at one glance.

The web application contributes to research and practice in several ways. First, the prototype provides coaches and managers with more lucid and conceptually grounded options for discovering and selecting appropriate strategic and tactical game plan decisions. By analyzing huge data sets, the prototype artifact enables users to receive a tangible and holistic overview of strengths and benefits of all available players for a soccer match. Thereby, our prototype can be used by coaches or managers to counteract particularly dangerous strategic situations and ensure the best possible outcome of the game from the outset. Figure 3 provides an overview of the application and evaluation process of our prototype: by aligning upcoming soccer game plans to the results from our prototype that provides BDA-based decisional support, new data will be generated, which will again serve as new data base for the prototype artifact. Thus, in the future, more data will be able to provide higher accuracy in the analysis of single players' performances.

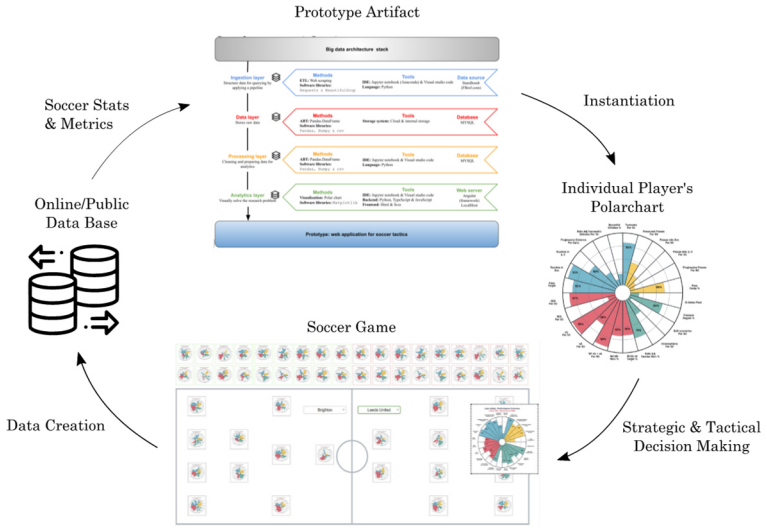

Fig. 3. Application and Effect Cycle of the Prototype Artifact

4 Conclusion and Outlook

The theoretical aim of this research is to contribute to the socio-technical challenge of how BDA can simplify the strategic tactical decision-making processes in soccer. As a result, our artifact can be viewed as a first step in bridging the gap between the complexity of BDA approaches and the need for a simple tool for strategic analysis in sports. We followed a design science research built-evaluate cycle and developed a

BDA-based web application, which aims at providing theoretically and conceptually grounded decisional guidance. Thus, a data-driven system based on polar chart templates is provided, which is able to process raw data from external sources (i.e., StatsBomb) in order to create holistic data visualizations identifying given prospects and tactical patterns according to preferences. Drawn from the consensus among the interviewed experts, the system has shown great potential in generalizing the strategic process of identifying tactical patterns. Subsequently, our prototype tool contributes to a deeper comprehension of big data's potential impact on strategic decisions in soccer tactics and for the domain of sports in general.

For future research, we plan on testing the artifact by observing a soccer club while utilizing the system for further proof-of-use and proof-of-value analyses. Furthermore, a consideration of the players' interaction as well as the influence of external variables like differing weather conditions could be included in the analysis. As BDA-driven data systems and applications expand, they will inevitably alter longstanding conceptions about decision making, competitive strategy formulation, management practices, and value creation in soccer.

References

1. Pappas, I.O., Mikalef, P., Giannakos, M.N., Krogstie, J., Lekakos, G.: Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies. Inf Syst E-Bus Manage. 16, 479–491 (2018).
2. 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., Lemmink, K.A.P.M.: Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. Europea Journal of Sport Science. 21, 481–496 (2021).
3. Memmert, D., Rein, R.: Match analysis, Big Data and tactics: current trends in elite soccer. German Journal of Sports Medicine. 2018, 65–72 (2018).
4. Rein, R., Memmert, D.: Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. SpringerPlus. 5, 1410 (2016).
5. Sonnenberg, C., Brocke, J. vom: Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research. In: Peffers, K., Rothenberger, M., and Kuechler, B. (eds.) Design Science Research in Information Systems. Advances in Theory and Practice. pp. 381–397. Springer Berlin Heidelberg (2012).
6. Ivanov, T., Singhal, R.: ABench: Big Data Architecture Stack Benchmark. In: Companion of the 2018 ACM/SPEC International Conference on Performance Engineering. pp. 13–16. ACM, Berlin Germany (2018).
7. Erraissi, A., Belangour, A.: Data sources and ingestion big data layers: meta-modeling of key concepts and features. Int Journal of Engineering & Technology. 7, 3607–3612 (2018).
8. StatsBomb: StatsBomb | Football Like Never Before, https://statsbomb.com/.
9. Mitchell, R.E.: Web scraping with Python: collecting more data from the modern web. O'Reilly Media, Sebastopol, CA (2018).
10. Palanivel, K.: Modern Network Analytics Architecture Stack to Enterprise Networks. Int Journal for Research in Applied Science and Engineering Technology. 7, 2634 (2019).