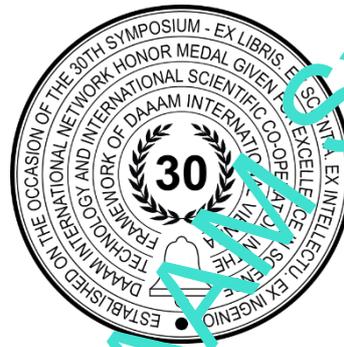# MACHINE LEARNING FOR DATA ANALYSIS IN FOOTBALL: A SURVEY OF METHODS AND PROBLEMS

Saša Tokić, Ante Panjkota & Maja Matetić

**Abstract**

Machine learning is growing exponentially, and its applications are gaining more traction in the sports analysis community in recent years. The application of machine learning methods on spatiotemporal data in sports like football is getting attention from football clubs, academics, and amateur analysts and is the focus of this survey.
This survey analyses and identifies current trends in research papers and literature to determine current and future applications in football analytics using spatiotemporal data.

**Keywords:** spatiotemporal data; sports analytics; event data; deep neural networks, machine learning

## 1. Introduction

Sports analytics in sports clubs has been evolving ever since "Moneyball" [1] showed how applied statistics could be used to select less know but valuable baseball players. While many sports, including football, have had rich data for decades until recently, spatiotemporal data was unavailable, and the advancement of technical capabilities enabled the evolution of available data. While so-called match sheet data certainly has its' value and is mainly understood by experts and laymen alike, spatiotemporal data opens new possibilities for clubs and scouts. Unfortunately, football clubs focus primarily on video analysis, which is highly time-consuming and introduces bias.

Spatiotemporal event data helps in answering many practical questions of much interest in football clubs like:

- Calculate the probability of scoring a goal from a given situation?
- How to evaluate a particular pass?
- What are common tactics opponents use?
- Which players have similar playing styles?

This paper analysis and identifies current trends in research papers and literature to determine current and future applications in sports analytics using machine learning and spatiotemporal data.

The goal of this research was to research the current state of data types, data availability, and research methods and applications of machine learning in sports analytics, focusing on applied research in football. Before the literature overview was conducted, we proposed the following research questions:

- Research question 1: How to categorize current research in machine learning in football analytics
- Research question 2: What are common methods of machine learning in football analytics
- Research question 3: What are possible further research directions

To answer our research questions, we've conducted a comprehensive literature overview using the Systematic Literature Review approach [2] to gather recent literature. The main goal was to classify recent research and analyse then categorize it by applying the content analysis method. The result of the analysis showed that the literature could be classified broadly as applying machine learning to spatiotemporal data in football to:

- Evaluate individual players
- Evaluate passes/actions
- Evaluate and classify teams
- Game result prediction

This paper is organized as follows: first, we introduce the domain of sports analytics and systemize the types of data and their availability; second, we provide an overview of recent research in the field of machine learning and spatiotemporal data in football; finally, we give a conclusion of the current state and suggest further research directions possible.

## 2. Data types and availability

Data suitable for football (soccer in the USA) analytics can be divided into three categories:

- Match sheet data
- Event data
- Tracking data

All types of data are obtained from specialized companies or websites. For example, service providers like FBref [3] usually provide match sheet data without fees, while event and tracking data are only available under a proprietary license. Some existing open datasets of event data are provided by providers like Wyscout [4] and Statsbomb [5]. Tracking data providers include Opta [6], Signality [7], SecondSpectrum [8], Metrica [9], and others.

Match sheet data provides a high-level statistic of the game, player, or club. In general, this data is freely accessible, and some websites even encourage the public to "scrape" the data from them.

Many sources like FBref [3] provide additional statistical data, such as adding Expected Goals and Expected Assists. Still, this kind of data doesn't help answer complex questions club stakeholders might have.

Event data is a sparser type of data, like tracking data; it is generated from video by human annotators and consists of current player and ball positions. Data points occur after a particular event like a pass, goal, foul or other.

Although this data also is usually not freely available, some available datasets are appropriate for research. The most significant difference between event and tracking data is that tracking data tracks the position of all players on the pitch. In contrast, event data only records the event, disregarding other players' positions.

Pappalardo describes the biggest available dataset to date in [10]. Data covers the seven most significant European leagues, World Cup 2018. and European cup 2016. Event data consists of events like pass, foul, and others with subtypes like cross-pass.

Tracking (spatiotemporal) data represents a type of data that consists of both time and space with unique properties regarding the modeling of spatiotemporal relations [11]. For example, in sports, there are usually 10-30 data points per second (10-30Hz) representing players' current positions on the pitch plus the ball. Tracking data can be obtained in multiple ways – static cameras, commercial video broadcasts, and GPS devices. However, due to its high commercial value, it is challenging for researchers to obtain this data type. Therefore, this data type is usually reserved for analysts in clubs with access to their league data.

Khaustov and Morgovoy [12] propose "a rule-based algorithm for identifying several basic types of events in soccer, including ball possession, successful and unsuccessful passes, and shots on goal." Direct potential benefits of applying their algorithm to tracking data could enhance additional event information usually unavailable as part of tracking data., p.

3. **Research**

The research questions in the introduction are self-evident, bearing in mind that the research is broadly interested in the state-of-art machine learning applications in spatiotemporal data analysis in football.

Systematic Literature Review guidelines [2] are used to answer these questions. In that sense, a literature search was conducted extensively in the IEEE Xplore Digital Library (IEEE), Scopus database, Web of Science database (WoS), ScienceDirect, and Google Scholar. Furthermore, due to the specificity of the research topic, search terms include a specific combination of targeted keywords, e.g., "Machine learning" AND "spatiotemporal" AND "football" OR "soccer" and many others.

In the next step, we created a table where we classified all the research papers by data type used (event or tracking), machine learning method, practical application, and broad categories that could group relevant papers. Some research papers were included in more than one category.

The research categories that were most apparent from the literature overview of 45 most recent or most important research papers with a corresponding number of papers are as follows:

- Evaluate individual players (13)
- Evaluate passes/actions (12)
- Evaluate and classify teams (11)
- Game result prediction (8)

4. **Results**

After the literature overview it became evident that the application of machine learning models to football analysis varies from so-called classical methods like a k-nearest neighbours to complex neural networks. One of the most compelling papers which provide football-applicable CNN architecture [13] is "Soccermap" [14] shown in Figure 1
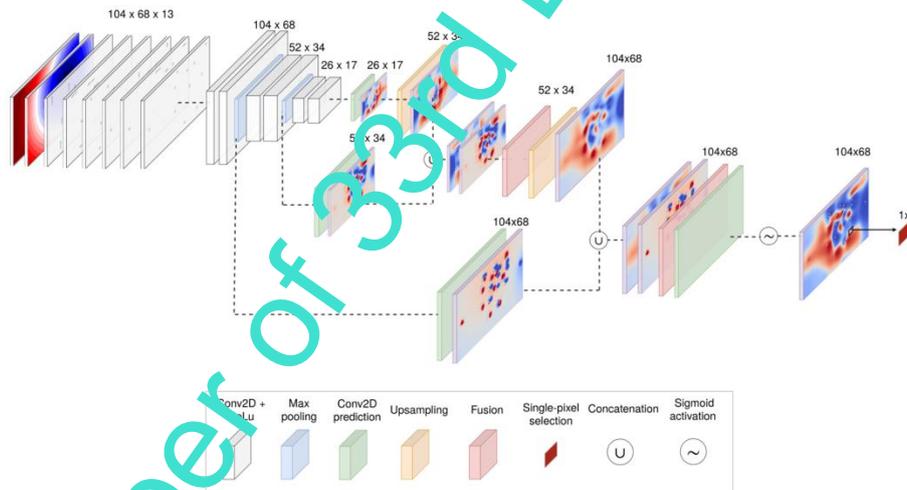


Fig. 1. Soccermap CNN, [14]

This architecture demonstrates that it is possible to apply modern machine learning approaches to tracking data with practical applications in football. By changing only, the output function, this network can calculate the probability surfaces of potential passes, estimate the pass-selection likelihood, and predict the expected value of the pass.

In Table 1, we've organized the most recent and influential research in machine learning in football, mainly focusing on spatiotemporal data. Correspondingly, we provide an overview of data types, methods, practical applications, and identified tasks.

| Reference | Data | Method | Application | Task |
|---|---|---|---|---|
| Cintia et al. [15] | Event | k-means, autoencoder | capturing and analysing the playing style of players, teams, and coaches in an automatic way | Evaluate and classify teams |
| Fernández et al. [16] | Tracking | Markov decision process | ability to evaluate the impact of observed and potential actions, both visually and analytically, expected possession value | Evaluate passes/actions |
| Hyeonah et al. [17] | Event | Convolutional Autoencoder | characterizing player's styles | Evaluate passes/actions |
| Kim et al [18] | Tracking | CNN, embeddings | Detecting players and their styles | Evaluate individual players |
| Lee and Jung [19] | Match sheet | DNN | predicting soccer tactics | Evaluate and classify teams |
| Malamatinos et al. [20] | Match sheet | k-Nearest Neighbour (k-NN), LogitBoost (LB), Support Vector Machine (SVM), Random Forest (RF), and CatBoost (CB) | | Game result prediction |
| Kim et al. [21] | Tracking | graph based CPD | distinguishes tactically intended formation and role changes from temporary changes in soccer matches | Evaluate and classify teams |
| Clijmans et al. [22] | Event | Markov model | analysing playing style | Evaluate and classify teams |
| Goes et al. [23] | Tracking | KMeans | Classify attacks as successful | Evaluate passes/actions |
| Fernàndez et al. [14] | Tracking | CNN | probability surfaces of potential passes, the expected value of a pass | Evaluate passes/actions |
| Goes et al. [24] | Tracking | KMeans | defenders create space for attackers and strongly dependent on those attacks' success | Evaluate passes/actions |
| Raabe et al. [25] | Event | Graph neural networks | Analysing tactical patterns | Evaluate and classify teams |
| Verstraete et al. [26] | | CPD (canonical polyadic decomposition) [27] | | Evaluate individual players |
| Nunez and Dagnino [28] | | pitch control, expected possession value and expected goals in a weighted function | Google Research Football competition [29] | - |
| Liu et al. [30] | Event | Deep Reinforcement Learning, LSTM | developed a new metric called GIM, to evaluate teams | Evaluate and classify teams |
| Decroos et al.[31] | Event | non-negative matrix factorisation [32] | identifying players with a similar style | Evaluate individual players |
| Beal et al. [33] | Event | walks within graphs, mixed integer programming | forming teams with | Evaluate individual players |
| Pappalardo et al. [34] | Event | Linear Support Vector Classifier | role-aware player performance evaluation | Evaluate individual players |
| Groll et al. [35] | Event | random forests and Poisson ranking | | Game result prediction |
| Goes et al. [36] | Tracking | Principal component analysis [37] | evaluating pass value | Evaluate passes/actions |
| Dick et al. [38] | Tracking | deep reinforcement learning | valuations of multiple players positioning | Evaluate individual players |
| Yam [39] | Tracking | deep reinforcement learning | valuations of multiple players positioning | Evaluate individual players |
| Decroos et al. [40] | Event | Generalized Additive Model [41] | improvement to VAEP model [42] | Evaluate passes/actions |
| Bransen et al. [43] | Event | distance-weighted k-nearest neighbours search | a new metric that aims to measure players' contribution in creating goal-scoring chances | Evaluate individual players |
| Zambom-Ferraresi et al. [44] | Tracking | Bayesian model averaging [45] | discover determinants of sports performance | Evaluate passes/actions |
| Steiner et al. [46] | Tracking | binary logistic regressions | estimate the | Evaluate passes/actions |

| | | | effects of contextual features on passing decisions | |
|---|---|---|---|---|
| Pappalardo et al. [47] | Tracking | OLS regression and logit classification | finds team ranking in the future season by using data from previous seasons | Evaluate and classify teams |
| McHale et al. [48] | Tracking | probability of a successful pass and network centrality measures | help trainers and scouts identify vital players in either opposition teams when recruiting new talents | Evaluate individual players |
| Giancola et al. [49] | - | CNN, ResNet-152 | detecting events in football broadcast videos | - |
| Decroos 26/10/2022 19:03:00et al. [49] | Event | clustering | discover tactics of football teams | Evaluate and classify teams |
| Decroos et al. [42] | Event | Clustering, exponential-decay-based | find top performing players in a league or a particular match | Evaluate individual players |
| Steiner 26/10/2022 19:03:00et al. [46] | Event | regression model | whom the player is most likely to pass the ball | Evaluate passes/actions |
| Horton et al. [50] | Tracking | computational geometry features fed into different classifiers (MLR, SVM, RUSBoost | classify the quality of passes in football | Evaluate passes/actions |
| Brooks et al. [51] | Event | L2-regularized Support Vector Machine (SVM) model | rank players based on the value of their passes | Evaluate individual players / Evaluate passes/actions |
| Brooks et al. [52] | Event | KNN | create unique team identification | Evaluate and classify teams |
| McHale et al. [53] | Event | mixed-effects model | model for identifying the ability of football players to score goals | Evaluate individual players |
| Bialkowski et al. [54] | Tracking | minimum entropy data partitioning and expectation-maximization (EM) algorithm | analysing both individual players and teams | Evaluate and classify teams / Evaluate individual players |
| Bialkowski et al. [55] | Tracking | LDA/ k-NN | identifying teams from spatiotemporal tracking data | Evaluate and classify teams |
| Tewari et al. [56] | | XGboost, SVM, Logistic regression | - | Game result prediction |
| Baboota et al. [57] | | Gaussian naive Bayes, SVM, Random forest, and gradient boost | | Game result prediction |
| Razali et al. [58] | | Bayesian networks | | Game result prediction |
| Danisik et al. [59] | | LSTM regression model | | Game result prediction |
| Cho et al. [60] | | gradient boosting | | Game result prediction |
| Ulmer et al. [61] | | Naive Bayes, Hidden Markov Model, Multinomial Naive Bayes, RBF SVM, Random forest, Linear SVM, One vs. All SGD | | Game result prediction |

Table 1. Literature overview, Source: authors contribution

## 5. Conclusion

This review aimed to evaluate the current state of machine learning in sports analytics, particularly interested in applications on data analytics in football. While the field might seem like a niche category, the number of research papers and non-academic content was pretty significant, so we've decided to focus on the most recent studies most representative of the current state-of-the-affairs. Besides many possible research directions from the literature, this review detects four main research tasks.

These tasks are:

- players evaluation
- passes and actions evaluation
- team evaluation
- game result prediction

It is no surprise that most tasks have practical applications that are of interest to football club stakeholders. Such claims include scouting, which stems from player evaluation, overall team formation improvement by combining player evaluation and team evaluation, and team success expectations by combining all four tasks are just a few of many.

The coaching staff has many benefits, including tactical preparations of matches from passes/actions evaluation, players evaluation, and opposing team evaluation. In addition, game result predictions are also of interest to sports betting as in academic research.

On the other hand, conducted results show that methods in this area span from so-called classical machine learning to deep learning, which has been the dominant method in recent times. Also, data types are closely related to the method used; in that sense, CNN models have shown the most promising research direction.

We intentionally omit the area of simulation environments methods from this survey because it is the emerging approach to addressing problems. Among them, we would like to point out possible research directions in simulated multi-agent environments like Google Research Football as a method for coping with defined tasks.

## 6. References

[1] Lewis, M. (2004). Moneyball: the art of winning an unfair game, Norton, 978-0-393-05765-2, New York, NY
[2] Kitchenham, B. & Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering, Keele, Staffs, UK, School of Computer Science and Mathematics, Keele University
[3] https://fbref.com/en/ (2020). FBref - Football Statistics and History, FBref.com, Accessed on: 2020-10-9
[4] https://wyscout.com/ (2020). Wyscout - Football Professional Videos and Data Platform, Wyscout, Accessed on: 2020-10-10
[5] http://statsbomb.com/ (2020). StatsBomb | Football Like Never Before, StatsBomb, Accessed on: 2020-10-10
[6] https://www.optasports.com (2020). Opta, Opta Sports, Accessed on: 2020-10-11
[7] https://www.signality.com/ (2020). Signality, Accessed on: 2020-11-10
[8] https://www.secondspectrum.com/index.html (2020). Second Spectrum, Accessed on: 2020-11-10
[9] https://github.com/metrica-sports/sample-data (2020). Metrica Sports Sample Data, Accessed on: 2020-10-10
[10] 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, Vol. 6, No. 1, pp. 236, 2052-4463, https://doi.org/10.1038/s41597-019-0247-7
[11] Salibekyan, S. & Pakulov, P. (2016). Database Architecture for Specifying and Modeling Spatio-Temporal Relations, In B. Katalinic (Ed.), DAAAM Proceedings, (1st ed., Vol. 1, pp. 0589–0598) DAAAM International Vienna, 978-3-902734-07-5 https://doi.org/10.2507/26th.daaam.proceedings.080
[12] Khaustov, V. & Mozgovoy, M. (2020). Recognizing Events in Spatiotemporal Soccer Data, Applied Sciences, Vol. 10, No. 22, pp. 8046, 2076-3417, https://doi.org/10.3390/app10228046
[13] LeCun, Y.; Boser, B.; Denker, J. S.; Henderson, D.; Howard, R. E.; Hubbard, W. & Jackel, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation, Vol. 1, No. 4, pp. 541–551, 0899-7667, 1530-888X, https://doi.org/10.1162/neco.1989.1.4.541
[14] Fernández, J. & Bornn, L. (2020). SoccerMap: A Deep Learning Architecture for Visually-Interpretable Analysis in Soccer, Presented at the The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML-PKDD, Virtual
[15] Cintia, P. & Pappalardo, L. (2021). Coach2vec: autoencoding the playing style of soccer coaches, arXiv https://doi.org/10.48550/arXiv.2106.15444

[16] Fernández, J.; Bornn, L. & Cervone, D. (2021). A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions, Machine Learning, Vol. 110, No. 6, pp. 1389–1427, 1573-0565, https://doi.org/10.1007/s10994-021-05989-6

[17] Cho, H.; Ryu, H. & Song, M. (2022). Pass2vec: Analyzing soccer players' passing style using deep learning, International Journal of Sports Science & Coaching, Vol. 17, No. 2, pp. 355–365, 1747-9541, https://doi.org/10.1177/17479541211033078

[18] Kim, H.; Kim, J.; Chung, D.; Lee, J.; Yoon, J. & Ko, S.-K. (2022). 6MapNet: Representing Soccer Players from Tracking Data by a Triplet Network, In U. Brefeld, J. Davis, J. Van Haaren, & A. Zimmermann (Eds.), Machine Learning and Data Mining for Sports Analytics, (pp. 3–14) Springer International Publishing, 978-3-031-02044-5, Cham https://doi.org/10.1007/978-3-031-02044-5_1

[19] Lee, G. J. & Jung, J. J. (2022). DNN-based multi-output model for predicting soccer team tactics, PeerJ Computer Science, Vol. 8, No. 1, pp. e853, 2376-5992, https://doi.org/10.7717/peerj-cs.853

[20] Malamatinos, M.-C.; Vrochidou, E. & Papakostas, G. A. (2022). On Predicting Soccer Outcomes in the Greek League Using Machine Learning, Computers, Vol. 11, No. 9, pp. 133, 2073-431X, https://doi.org/10.3390/computers11090133

[21] Kim, H.; Kim, B.; Chung, D.; Yoon, J. & Ko, S.-K. (2022). SoccerCPD: Formation and Role Change-Point Detection in Soccer Matches Using Spatiotemporal Tracking Data, In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, (pp. 3146–3156) Association for Computing Machinery, 978-1-4503-9385-0, New York, NY, USA https://doi.org/10.1145/3534678.3539150

[22] Clijmans, J.; Roy, M. V. & Davis, J. (2022). Looking Beyond the Past: Analyzing the Intrinsic Playing Style of Soccer Teams, In ECML PKDD 2022,Presented at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases

[23] Goes, F. R.; Brink, M. S.; Elferink-Gemser, M. T.; Kempe, M. & Lemmink, K. A. P. M. (2021). The tactics of successful attacks in professional association football: large-scale spatiotemporal analysis of dynamic subgroups using position tracking data, Journal of Sports Sciences, Vol. 39, No. 5, pp. 523–532, 0264-0414, https://doi.org/10.1080/02640414.2020.1834689

[24] Goes, F. R.; Brink, M. S.; Elferink-Gemser, M. T.; Kempe, M. & Lemmink, K. A. P. M. (2020). The tactics of successful attacks in professional association football: large-scale spatiotemporal analysis of dynamic subgroups using position tracking data, Journal of Sports Sciences, Vol. 0, No. 0, pp. 1–10, 0264-0414, https://doi.org/10.1080/02640414.2020.1834689

[25] Raabe, D.; Nabben, R. & Memmert, D. (2022). Graph representations for the analysis of multi-agent spatiotemporal sports data, Applied Intelligence, 1573-7497, https://doi.org/10.1007/s10489-022-03631-z

[26] Verstraete, K.; Decroos, T.; Coussement, B.; Vannieuwenhoven, N. & Davis, J. (2020). Analyzing Soccer Players' Skill Ratings Over Time Using Tensor-Based Methods, In P. Cellier & K. Driessens (Eds.), Machine Learning and Knowledge Discovery in Databases, (Vol. 1168, pp. 225–234) Springer International Publishing, 978-3-030-43886-9, Cham https://doi.org/10.1007/978-3-030-43887-6_17

[27] Sorber, L.; Van Barel, M. & De Lathauwer, L. (2013). Optimization-Based Algorithms for Tensor Decompositions: Canonical Polyadic Decomposition, Decomposition in Rank-$(L_r,L_r,1)$ Terms, and a New Generalization, SIAM Journal on Optimization, Vol. 23, No. 2, pp. 695–720, 1052-6234, 1095-7189, https://doi.org/10.1137/120868323

[28] Nunez, J. C. & Dagnino, B. (2020). Exploring the application of soccer mathematical models to game generation on a simulated environment, , pp. 10,

[29] https://kaggle.com/c/google-football (2020). Google Research Football with Manchester City F.C., Accessed on: 2020-10-4

[30] Liu, G.; Luo, Y.; Schulte, O. & Kharrat, T. (2020). Deep soccer analytics: learning an action-value function for evaluating soccer players, Data Mining and Knowledge Discovery, Vol. 34, No. 5, pp. 1531–1559, 1384-5810, 1573-756X, https://doi.org/10.1007/s10618-020-00705-9

[31] Decroos, T. & Davis, J. (2020). Player Vectors: Characterizing Soccer Players' Playing Style from Match Event Streams, In U. Brefeld, E. Fromont, A. Hotho, A. Knobbe, M. Maathuis, & C. Robardet (Eds.), Machine Learning and Knowledge Discovery in Databases, (Vol. 11908, pp. 569–584) Springer International Publishing, 978-3-030-46132-4, Cham https://doi.org/10.1007/978-3-030-46133-1_34

[32] Lee, D. D. & Seung, H. S. (2001). Algorithms for Non-negative Matrix Factorization, (p. 7)Presented at the NIPS

[33] Beal, R.; Changder, N.; Norman, T. & Ramchurn, S. (2020). Learning the Value of Teamwork to Form Efficient Teams, Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 05, pp. 7063–7070, 2374-3468, 2159-5399, https://doi.org/10.1609/aaai.v34i05.6192

[34] Pappalardo, L. & Cintia, P. (2018). Quantifying the relation between performance and success in soccer, Advances in Complex Systems, Vol. 21, No. 03n04, pp. 1750014, 0219-5259, 1793-6802, https://doi.org/10.1142/S021952591750014X

[35] Groll, A.; Ley, C.; Schauberger, G. & Van Eetvelde, H. (2019). A hybrid random forest to predict soccer matches in international tournaments, Journal of Quantitative Analysis in Sports, Vol. 15, No. 4, pp. 271–287, 1559-0410, 2194-6388, https://doi.org/10.1515/jqas-2018-0060

[36] Goes, F. R.; Kempe, M.; Meerhoff, L. A. & Lemmink, K. A. P. M. (2019). Not Every Pass Can Be an Assist: A Data-Driven Model to Measure Pass Effectiveness in Professional Soccer Matches, Big Data, Vol. 7, No. 1, pp. 57–70, 2167-6461, 2167-647X, https://doi.org/10.1089/big.2018.0067

[37] Wold, S.; Esbensen, K. & Geladi, P. (n.d.). Principal Component Analysis, , pp. 16,

[38] Dick, U. & Brefeld, U. (2019). Learning to Rate Player Positioning in Soccer, Big Data, Vol. 7, No. 1, pp. 71–82, 2167-6461, 2167-647X, https://doi.org/10.1089/big.2018.0054

[39] https://statsbomb.com/2019/02/attacking-contributions-markov-models-for-football/ (2019). Attacking Contributions: Markov Models for Football, StatsBomb, Accessed on: 2020-10-4

[40] Decroos, T. (2020). Soccer Analytics Meets Artificial Intelligence: Learning Value and Style from Soccer Event Stream Data, Ph.D. Dissertation, KU Leuven – Faculty of Engineering Science, Heverlee,

[41] Hastie, T. & Tibshirani, R. (1986). Generalized Additive Models, Statistical Science, Vol. 1, No. 3, pp. 297–310, 0883-4237, https://doi.org/10.1214/ss/1177013604

[42] Decroos, T.; Bransen, L.; Van Haaren, J. & Davis, J. (2020). VAEP: An Objective Approach to Valuing On-the-Ball Actions in Soccer (Extended Abstract), In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, (pp. 4696–4700)Presented at the Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence {IJCAI-PRICAI-20}, International Joint Conferences on Artificial Intelligence Organization, 978-0-9992411-6-5, Yokohama, Japan https://doi.org/10.24963/ijcai.2020/648

[43] Bransen, L.; Van Haaren, J. & van de Velden, M. (2019). Measuring soccer players' contributions to chance creation by valuing their passes, Journal of Quantitative Analysis in Sports, Vol. 15, No. 2, pp. 97–116, 1559-0410, 2194-6388, https://doi.org/10.1515/jqas-2018-0020

[44] Zambom-Ferraresi, F.; Rios, V. & Lera-Lopez, F. (2018). Determinants of sport performance in European football: What can we learn from the data?, Decision Support Systems, Vol. 114, , pp. 18–28, 01679236, https://doi.org/10.1016/j.dss.2018.08.006

[45] Volinsky, C. T.; Raftery, A. E.; Madigan, D. & Hoeting, J. A. (1999). Bayesian model averaging: a tutorial, Statistical Science, Vol. 14, No. 4, pp. 382–417, 0883-4237, https://doi.org/10.1214/ss/1009212519

[46] Steiner, S.; Rauh, S.; Rumo, M.; Sonderegger, K. & Seiler, R. (2018). Using position data to estimate effects of contextual features on passing decisions in football, , pp. 9,

[47] Pappalardo, L.; Cintia, P.; Ferragina, P.; Massucco, E.; Pedreschi, D. & Giannotti, F. (2019). PlayeRank: Data-driven Performance Evaluation and Player Ranking in Soccer via a Machine Learning Approach, ACM Transactions on Intelligent Systems and Technology, Vol. 10, No. 5, pp. 1–27, 21576904, https://doi.org/10.1145/3343172

[48] McHale, I. G. & Relton, S. D. (2018). Identifying key players in soccer teams using network analysis and pass difficulty, European Journal of Operational Research, Vol. 268, No. 1, pp. 339–347, 03772217, https://doi.org/10.1016/j.ejor.2018.01.018

[49] Decroos, T. (2018). STARSS: A Spatio-Temporal Action Rating System for Soccer, , pp. 10,

[50] Horton, M.; Gudmundsson, J.; Chawla, S. & Josephan, J. (2017). Classification of Passes in Football Matches using Spatiotemporal Data, ACM Transactions on Spatial Algorithms and Systems, Vol. 3, No. 2, pp. 1–30, 23740353, https://doi.org/10.1145/3105576

[51] Brooks, J.; Kerr, M. & Guttag, J. (2016). Using machine learning to draw inferences from pass location data in soccer: Drawing Inferences from Pass Location Data Using Machine Learning, Statistical Analysis and Data Mining: The ASA Data Science Journal, Vol. 9, No. 5, pp. 338–349, 19321864, https://doi.org/10.1002/sam.11318

[52] Brooks, J.; Kerr, M. & Guttag, J. (2016). Developing a Data-Driven Player Ranking in Soccer Using Predictive Model Weights, In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, (pp. 49–55)Presented at the the 22nd ACM SIGKDD International Conference, ACM Press, 978-1-4503-4232-2, San Francisco, California, USA https://doi.org/10.1145/2939672.2939695

[53] McHale, I. G. & Szczepanski, L. (2014). A mixed effects model for identifying goal scoring ability of footballers, Journal of the Royal Statistical Society: Series A (Statistics in Society), Vol. 177, No. 2, pp. 397–417, 09641998, https://doi.org/10.1111/rssa.12015

[54] Bialkowski, A.; Lucey, P.; Carr, P.; Yue, Y.; Sridharan, S. & Matthews, I. (2014). Identifying Team Style in Soccer Using Formations Learned from Spatiotemporal Tracking Data, In 2014 IEEE International Conference on Data Mining Workshop, (pp. 9–14)Presented at the 2014 IEEE International Conference on Data Mining Workshop (ICDMW), IEEE, 978-1-4799-4274-9, Shenzhen, China https://doi.org/10.1109/ICDMW.2014.167

[55] Bialkowski, A.; Lucey, P.; Carr, P.; Yue, Y.; Sridharan, S. & Matthews, I. (2014). Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data, In 2014 IEEE International Conference on Data Mining, (pp. 725–730)Presented at the 2014 IEEE International Conference on Data Mining (ICDM), IEEE, 978-1-4799-4302-9, Shenzhen, China https://doi.org/10.1109/ICDM.2014.133

[56] Tiwari, A.; Parwani, T.; Phanse, A.; Sharma, A. & Shetty, A. (2019). Soccer Analytics using Machine Learning, International Journal of Computer Applications, Vol. 181, No. 50, pp. 54–56, 09758887, https://doi.org/10.5120/ijca2019918773

[57] Baboota, R. & Kaur, H. (2019). Predictive analysis and modelling football results using machine learning approach for English Premier League, International Journal of Forecasting, Vol. 35, No. 2, pp. 741–755, 01692070, https://doi.org/10.1016/j.ijforecast.2018.01.003

[58] Razali, N.; Mustapha, A.; Yatim, F. A. & Aziz, R. A. (2017). Predicting Football Matches Results using Bayesian Networks for English Premier League (EPL), https://doi.org/10.1088/1757-899X/226/1/012099

[59] Danisik, N.; Lacko, P. & Farkas, M. (2018). Football Match Prediction Using Players Attributes, In 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA), (pp. 201–206)Presented at the 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA), IEEE, 978-1-5386-5102-5, Kosice https://doi.org/10.1109/DISA.2018.8490613

[60] Cho, Y.; Yoon, J. & Lee, S. (2018). Using social network analysis and gradient boosting to develop a soccer win-lose prediction model, Eng. Appl. Artif. Intell., https://doi.org/10.1016/j.engappai.2018.04.010

[61] Ulmer, B. & Fernandez, M. (2014). Predicting Soccer Match Results in the English Premier League (p. )Stanford University Retrieved from https://www.semanticscholar.org/paper/Predicting-Soccer-Match-Results-in-the-English-Ulmer-Fernandez/d4dfbb3d1454416799b72004d3d88bae6bbd660a