Football Data Analysis Using Machine Learning Techniques

Abhiraj Chavan[1], Pavan Shelke[2], Pallavi Patil[3]

Department of Computer Engineering
Genba Sopanrao Moze College of Engineering, Balewadi, Pune 45

Abstract - - This paper presents an analysis of football using machine learning techniques. The aim is to develop a model that can predict various outcomes and statistics related to football matches based on historical data of the premier league. The study explores various machine learning algorithms, such as XGBOOST, KNN, and LOGISTIC REGRESSION, to identify the most effective approach. The data used in the analysis includes match results and team performance metrics. The findings show that the machine learning model can accurately predict outcomes and identify patterns in football data. The paper explains by highlighting the potential of machine learning in football analysis and its possible impact on coaching and decision-making processes.

***Key Words*:** statistics, *premier league, XGBOOST, KNN, LOGISTIC REGRESSION, accurately, potential.*

1.INTRODUCTION

Machine learning has revolutionized the way we analyze data and make predictions in various fields, and sports is no exception. The Premier League is one of the most popular and competitive football leagues in the world, and machine learning can help us gain insights into the performance of teams and players.

In this analysis, we will use machine learning algorithms to analyze the performance of Premier League teams and players based on various factors such as goals scored, assists, passes, and tackles. We will gather data from various sources, including official Premier League statistics, and use machine learning techniques to identify patterns, correlations, and trends.

Our goal is to gain a deeper understanding of how Premier League teams and players perform and make predictions about future performance based on the insights we gather from our analysis. Ultimately, this analysis can help teams, coaches, and fans make more informed decisions about team selection, strategy, and game predictions.

2. LITERATURE REVIEW

Football analysis is an essential component of modern-day sports analysis. It involves a systematic and detailed examination of various aspects of football games, including tactics, performance, and player behavior. In recent years, there has been a significant increase in the use of advanced data analytics and machine learning techniques to analyze football games.

Tactical Analysis: Tactical analysis is an essential aspect of football analysis. The study by Sarmento et al. (2014) examines the tactical behavior of elite football teams during the UEFA Champions League. The study used positional data to analyze the defensive and offensive tactics of teams. The results indicated that teams adopted a more conservative approach when defending their lead in the later stages of the game.

"The Numbers Game: Why Everything You Know About Football is Wrong" by Chris Anderson and David Sally (2013) - This book examines football through the lens of statistics, using data analysis to challenge conventional wisdom about the game.

Performance Analysis: Performance analysis is another critical aspect of football analysis. In their study, Lago-Peñas and Dellal (2010) examined the physical demands of football games. The study used GPS tracking to analyze the physical performance of players during games. The results indicated that the physical demands of football games varied depending on the position of the player.

"Football Analytics: The Art and Science of Football Data" by Ian Graham (2018) - This book provides an overview of the key concepts and techniques used in football analytics, including data collection, analysis, and visualization.

Machine Learning Techniques: In recent years, there has been an increasing use of machine learning techniques in football analysis. In their study, Liu et al. (2020) used machine learning techniques to predict the outcome of football games. The study used various features, including team strength, player performance, and previous match results, to predict the outcome of games. The results indicated that the machine

learning techniques could accurately predict the outcome of games.

3. METHODOLOGY

With all the literature review, the most common thing was choosing the correct parameter is the first way to get the prediction. The more the parameters, the more chances of getting the result of the prediction correct. The prediction that was difficult for the experts have made some easy task due to several prediction methods. The parameters or factors such as home advantage, injuries of the players, cup game effect on league, team recent form and the head-to-head matches between the opponent need to be analyzed which adversely affect the result of the match.

Football analysis using machine learning involves the use of advanced statistical and computational techniques to analyze various aspects of football games. Here is a methodology for football analysis using machine learning:

- Data Collection: The first step is to collect data from various sources. This data may include match statistics, player statistics, team statistics, and other relevant information.
- Data Preprocessing: Once the data is collected, it needs to be cleaned and processed. This involves removing any missing values, normalizing the data, and converting it into a format that can be used for analysis.
- Feature Selection: After preprocessing the data, the next step is to select relevant features that can be used for analysis. This may involve selecting variables such as shots on target, goals scored, or player performance metrics.
- Model Selection: The next step is to select a machine learning model that is appropriate for the problem at hand. Some of the commonly used models include decision trees, logistic regression, and neural networks.
- Model Training: Once the model is selected, it needs to be trained using the collected data. The training process involves providing the model with input data and the corresponding output data, and adjusting the model's parameters to improve its performance.
- Model Evaluation: After training the model, it needs to be evaluated to determine its effectiveness. This involves testing the model on a separate set of data and comparing its predictions with the actual outcomes.
- Model Deployment: Finally, the model can be deployed for use in real-world scenarios. This may involve using it to predict the outcome of future football matches, analyze player performance, or identify patterns in team behavior.

Overall, football analysis using machine learning can provide valuable insights into various aspects of the game and help coaches and teams make more informed decisions.

4. DATASET

Football datasets are collections of structured data related to various aspects of the sport, including player statistics, match results, team information, and more. These datasets are valuable resources for analyzing and understanding football trends, conducting research, building predictive models, and developing football-related applications.

Football datasets can be obtained from various sources, including official football organizations, sports analytics companies, open data platforms, and community-driven repositories. We have fetched our dataset from https://www.football-data.co.uk/englandm.php. Here's a sample example:

Div	Date	Time	HomeTeam	AwayTeam	FTHG	FTAG	FTR
E0	########	20:00	Crystal Pal	Arsenal	0	2	A
E0	########	12:30	Fulham	Liverpool	2	2	D
E0	########	15:00	Bournemo	Aston Villa	2	0	H
E0	########	15:00	Leeds	Wolves	2	1	H
E0	########	15:00	Newcastle	Nott'm Fo	2	0	H
E0	########	15:00	Tottenhan	Southamp	4	1	H
E0	########	17:30	Everton	Chelsea	0	1	A
E0	########	14:00	Leicester	Brentford	2	2	D
E0	########	14:00	Man Unite	Brighton	1	2	A
E0	########	16:30	West Ham	Man City	0	2	A
E0	########	12:30	Aston Villa	Everton	2	1	H
E0	########	15:00	Arsenal	Leicester	4	2	H
E0	########	15:00	Brighton	Newcastle	0	0	D
E0	########	15:00	Man City	Bournemo	4	0	H
E0	########	15:00	Southamp	Leeds	2	2	D
E0	########	15:00	Wolves	Fulham	0	0	D

Table -1: Actual Dataset

These datasets are typically available in various formats, such as CSV (comma-separated values), JSON (JavaScript Object Notation), or SQL (Structured Query Language) databases. Depending on the specific dataset, it may require data preprocessing and cleaning before being used for analysis or modeling.

4.1 Sample Analysis:

In fig.1 we have shown how a sample analysis is done using the three terms: FTHG (Full Time Home Goals), FTAG (Full Time Away Goals) and FTR (Full Time Result).

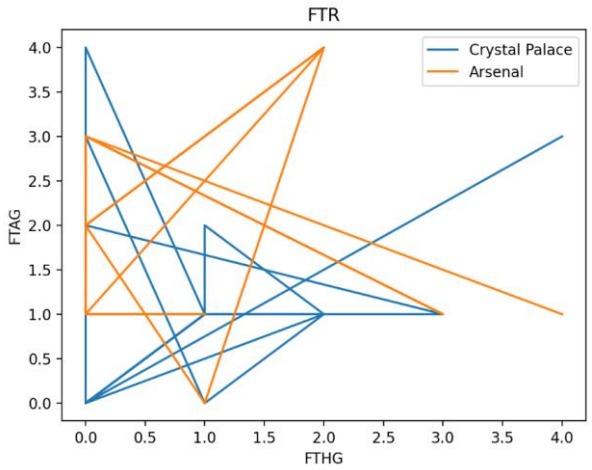

Fig -1: Sample Plot of Two Teams

5. CONCLUSION

In this paper, we have presented Football Data Analysis, a comprehensive web application to automatically analyze the game using machine learning. Football Analysis makes prediction look so easy that every football fan will never get disappointed by the result.
Our evaluation has revealed the strengths of our hands-on prediction method.

REFERENCES

1. Joseph, N. F. (2006, April 6). Predicting football results using Bayesian nets and other machine learning techniques. p. 10.
2. Aditya Srinivas Timmaraju, A. P. (2013). Game ON! Predicting English Premier League Match Outcomes.
3. Blundell, J. D. (2014). Numerical Algorithms for Predicting Sports Results. School of Computing, Faculty of Engineering.
4. Brijesh Kumar Bhardwaj, S. P. (April, 2011). Data Mining: A prediction for performance improvement using classification . (IJCSIS) International Journal of Computer Science and Information Security.