

Moringa School

Capstone Project Proposal

Soccer Prediction Model

Group 7

10 – 01- 2024

Table of Contents

Business Understanding	2
1.1 Background.....	2
1.2 Problem Statement	3
1.2.1 Challenges.....	3
1.2.2 Opportunity	3
1.3 Research Questions	4
1.3.1 General Research Questions:.....	4
1.3.2 Exploratory Data Analysis (EDA) Research Questions:	4
1.3.3 Additional Machine Learning Model Assessment Questions:.....	4
1.4 Project Objectives:	5
1.4.1 General Objective:	5
1.4.2 Specific Objectives:.....	5
2 Data Understanding:	6
2.1 Data Source:.....	6
2.2 Data Points and Descriptions:	6
2.3 Data Retrieval Process:	8
3 Success Criteria:	8
3.1 Model Performance:.....	8
3.1.1 Accuracy:	8
3.1.2 Consistency:	9
3.1.3 Benchmark Comparison:	9
3.2 User Experience:	9
3.2.1 User Satisfaction:	9
3.2.2 Accessibility:.....	9
3.3 Impact and Application:	9
3.3.1 Real-world Application	9
3.3.2 Adoption Rate:	9
3.4 Business Value:	10
3.4.1 Revenue Generation:	10
3.4.2 Market Recognition:.....	10
4 References	10

Business Understanding

1.1 Background

The betting industry in Kenya has undergone a transformative shift, evolving into a multi-billion-dollar industry. According to the Betting Control and Licensing Board of Kenya, the sector has grown exponentially, contributing significantly to the nation's economic development. The latest available data from the board indicates that the industry generated over Ksh 204 billion (approximately USD 2 billion) in revenue in the last fiscal year alone (Betting Control and Licensing Board - Kenya).

Within this burgeoning industry, soccer betting has emerged as the undisputed frontrunner, capturing the interest of millions of Kenyan sports enthusiasts. A survey conducted by GeoPoll, a mobile-based market research platform, revealed that over 60% of mobile gamers in Kenya actively engage in soccer betting, making it the dominant betting category (GeoPoll).

The allure of soccer betting lies in its dynamic and unpredictable nature, where punters navigate through a plethora of betting options. From predicting the winner of a match to anticipating the total goals scored, the process involves a meticulous analysis of team and player statistics, historical performances, and real-time match dynamics. Kenyan sports enthusiasts, fueled by their passion for the game, seek to translate their knowledge into profitable betting decisions (Daily Nation).

Recognizing the demand for more informed and accurate betting decisions, the intersection of technology and sports betting has witnessed a surge in predictive modeling. The Kenya Gazette, an official government publication, acknowledges the potential of predictive models in providing valuable insights to bettors, helping them make data-driven decisions to enhance their success rate (Kenya Gazette).

This research project focuses on the development of a simplified soccer match outcome prediction system for the English Premier League. The envisioned product is a user-friendly system where a client inputs a match ID, and the system, powered by machine learning, provides a prediction for the match outcome. Leveraging advanced algorithms, the system aims to offer straightforward and reliable match predictions based on historical data and relevant match features.

Amid the remarkable growth in the Kenyan betting industry, there exists a research gap in the application of machine learning to provide simplified match predictions for soccer enthusiasts. This project aims to address this gap by introducing cutting-edge data science techniques into the realm of Kenyan soccer betting, offering users a practical and accessible tool to enhance their betting experience.

1.2 Problem Statement

The surge in popularity of sports betting, particularly in the vibrant betting industry of Kenya, has prompted an increasing demand for accurate and data-driven predictions, especially in the context of soccer matches. Despite the availability of vast amounts of football-related data, there is a notable gap in providing simplified and user-friendly prediction systems for Kenyan soccer enthusiasts.

1.2.1 Challenges

- i. Existing machine learning predictive models often lack accessibility for the average user, requiring a level of expertise in data interpretation.
- ii. Additionally, current betting prediction platforms may not utilize advanced machine learning techniques to offer precise and tailored predictions.

1.2.2 Opportunity

The opportunity lies in developing a simplified and user-friendly soccer match outcome prediction system for the English Premier League. This system aims to empower users with

accurate match predictions, leveraging machine learning algorithms to process historical data and relevant match features.

1.3 Research Questions

1.3.1 General Research Questions:

- i. How can machine learning algorithms be effectively employed to predict soccer match outcomes?
- ii. What are the key features and historical data points that significantly influence match predictions?
- iii. How can a simplified prediction system be designed to cater to the needs of Kenyan soccer betting enthusiasts?

1.3.2 Exploratory Data Analysis (EDA) Research Questions:

- i. What are the distribution patterns of key performance metrics, such as team ranks, points, and goal differentials, across the English Premier League teams?
- ii. Are there any correlations or trends between a team's historical performance metrics and its current rank in the league standings?
- iii. How do teams' performance metrics vary when playing at home versus playing away?
- iv. Can patterns or trends be identified in the outcomes of matches based on the results of the last five games' results for each team?
- v. What role does goal difference play in determining the outcome of matches, and how does it correlate with the teams' overall performance?

1.3.3 Additional Machine Learning Model Assessment Questions:

- i. How sensitive is the prediction model to changes in input features, and which features contribute most significantly to prediction outcomes?
- ii. What impact do changes in the training dataset size and composition have on the model's predictive performance?

1.4 Project Objectives:

1.4.1 General Objective:

1.4.1.1 To Develop a Soccer Match Outcome Prediction System:

Create a robust and user-friendly machine learning-based system capable of predicting the outcome of English Premier League soccer matches, catering specifically to the preferences and needs of Kenyan soccer betting enthusiasts.

1.4.2 Specific Objectives:

1.4.2.1 To Implement Advanced Machine Learning Algorithms for Soccer Match Prediction:

Integrate sophisticated machine learning algorithms, such as ensemble methods or deep learning models, to enhance the accuracy and predictive power of the soccer match outcome prediction system.

1.4.2.2 To Conduct Feature Selection for Enhanced Predictive Insights:

Conduct a comprehensive analysis to identify and select key features and historical data points that significantly influence soccer match outcomes, ensuring the inclusion of the most impactful variables in the prediction model.

1.4.2.3 To Design a User-Friendly Interface for Soccer Match Predictions:

Develop an intuitive and user-friendly interface for the prediction system, optimizing accessibility for users with varying technical backgrounds and ensuring a seamless experience in obtaining match predictions.

1.4.2.4 To Iteratively Optimize the Prediction Model for Accuracy:

Implement an iterative optimization process to continuously refine and improve the prediction model, incorporating feedback and adjusting parameters to achieve the highest possible accuracy in predicting English Premier League soccer match outcomes.

1.4.2.5 To Validate Predictions Against Historical Match Data:

Validate the prediction system by comparing its predictions against actual outcomes using a historical dataset of English Premier League matches, establishing the reliability and effectiveness of the developed system.

2 Data Understanding:

2.1 Data Source:

The primary data source for this project will be the Football Data API (<https://footystats.org/api>), providing comprehensive information on English Premier League teams, matches, and performance metrics.

2.2 Data Points and Descriptions:

Data Point	Description
Team Rank	The current rank of the team within the English Premier League standings.
Team Name and Crest	The official name and emblem of the team for identification.
Played Matches	The total number of matches played by the team in the current season.
Wins	The total number of matches won by the team.
Losses	The total number of matches lost by the team.

Draws	The total number of matches drawn by the team.
Points	The total points earned by the team based on wins and draws.
Last Five Games Results	The outcomes (win, lose, draw) of the team's last five matches.
Goal Difference	The numerical difference between goals scored and goals conceded.
Differential	A calculated metric representing the difference between wins and losses.
Goals For	The total number of goals scored by the team.
Goals Against	The total number of goals conceded by the team.
Win Percentage	The percentage of matches won by the team.
Won in Group	The number of matches won by the team when playing in a group.
Lost in Group	The number of matches lost by the team when playing in a group.

Win Percentage in Groups	The percentage of matches won by the team when playing in a group.
Won at Home	The number of matches won by the team when playing at their home stadium.
Won Away	The number of matches won by the team when playing away.
Lost at Home	The number of matches lost by the team when playing at their home stadium.
Lost Away	The number of matches lost by the team when playing away.

2.3 Data Retrieval Process:

- i. Utilize the Football Data API to retrieve real-time and historical data for English Premier League teams.
- ii. Implement requests to obtain specific metrics outlined above for each team, considering relevant time frames for historical performance.

3 Success Criteria:

3.1 Model Performance:

3.1.1 Accuracy:

- Achieve a prediction accuracy of at least 70% in determining the correct outcome (win, lose, draw) of English Premier League soccer matches.

3.1.2 Consistency:

- Demonstrate consistency in predictions across multiple test datasets, indicating robust and reliable model performance.

3.1.3 Benchmark Comparison:

- Outperform or match the performance of existing benchmark models or prediction systems in the context of English Premier League match predictions.

3.2 User Experience:

3.2.1 User Satisfaction:

Conduct user feedback surveys to gauge user satisfaction with the prediction system, with a target satisfaction score of 80% or above.

3.2.2 Accessibility:

Ensure the system's user interface is user-friendly and accessible to individuals with varying levels of technical expertise.

3.3 Impact and Application:

3.3.1 Real-world Application

- Demonstrate the practical application of the prediction system by showcasing successful predictions on upcoming matches and assessing its impact on user betting decisions.

3.3.2 Adoption Rate:

- Achieve a significant adoption rate among Kenyan soccer betting enthusiasts, with a goal of at least 50,000 active users within the first six months of deployment.

3.4 Business Value:

3.4.1 Revenue Generation:

If applicable, generate revenue through partnerships, subscriptions, or user engagement features, aiming for a positive return on investment within the project's first year.

3.4.2 Market Recognition:

Gain recognition and positive reviews within the sports betting industry and among relevant stakeholders, positioning the prediction system as a valuable tool in the Kenyan market.

4 References

1. Betting Control and Licensing Board - Kenya. (n.d.). Retrieved from <https://bclb.co.ke/>
2. Daily Nation. (n.d.). Retrieved from <https://www.nation.co.ke/kenya/sports/-/1090/1090/-/h96psf/-/index.html>
3. GeoPoll. (n.d.). The Rise of Mobile Gaming in Africa. Retrieved from <https://www.geopoll.com/blog/the-rise-of-mobile-gaming-in-africa/>
4. Kenya Gazette. (n.d.). Official government publication. Retrieved from <https://www.kenyagazette.co.ke/>