



PROJECT REPORT
ON
English Premier League Match Predictor

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENT FOR SEMESTER VI OF

T.E.(Information Technology)

SUBMITTED BY

Mr. Shaun D'souza (IT5B094)

Mr. Abhinav Haridas (IT5B081)

Mr. Himnish Israni (IT5B099)

Mr. Kaushal Poojary (IT5B125)

UNDER THE GUIDANCE OF

Mrs. Kajal Jewani

DEPARTMENT OF INFORMATION TECHNOLOGY
V.E.S. INSTITUTE OF TECHNOLOGY
2022-23

Certificate

This is to certify that project entitled

”English Premier League Match Predictor”

Group Members Names

Mr. Shaun D’souza (Roll No. 14)
Mr. Abhinav Haridas (Roll No. 20)
Mr. Himnish Israni (Roll No. 22)
Mr. Kaushal Poojary (Roll No. 55)

In partial fulfillment of degree of BE. (Sem VI) in Information Technology for Project is approved.

Mrs. Kajal Jewani

External Examiner

Dr.(Mrs.)Shalu Chopra
H.O.D

Dr.(Mrs.)J.M.Nair
Principal

Date: 6/4/2023
Place: VESIT, Chembur

College Seal

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)

Shaun D'souza (Roll No.14)

(Signature)

Abhinav Haridas (Roll No.20)

(Signature)

Himnish Israni (Roll No.22)

(Signature)

Kaushal Poojary (Roll No.55)

ACKNOWLEDGEMENT

The project report on "English Premier League Match Predictor" is the outcome of the guidance, moral support and devotion bestowed on our group throughout our work. For this we acknowledge and express our profound sense of gratitude to everybody who has been the source of inspiration throughout project preparation. First and foremost we offer our sincere phrases of thanks and innate humility to "HOD Dr.(Mrs.)Shalu Chopra H.O.D", "Deputy HOD Mr. Manoj Sabnis", "Project guide Mrs. Kajal Jewani" for providing the valuable inputs and the consistent guidance and support provided by them. We can say in words that we must at outset tender our intimacy for receipt of affectionate care to Vivekanand Education Society's Institute of Technology for providing such a stimulating atmosphere and conducive work environment.

Abstract

This project aims to create an English premier league match predictor using machine learning techniques in Python. The predictor will use historical match data to make predictions about the outcomes of future matches. The project will explore various machine learning algorithms and techniques, such as regression, classification, and clustering, to find the best predictors of football match results. Football is a popular sport worldwide, with millions of fans following their favorite teams and players. Predicting the outcome of a football match is a challenging task due to the many factors that can influence the result. Machine learning techniques offer a promising approach to football match prediction, as they can analyze large amounts of data and find patterns that may not be immediately apparent to human analysts.

Contents

1	Introduction	1
1.1	Introduction	1
1.2	Aim and Objectives	1
1.3	Motivation for the Work	2
1.4	Scope of Project	2
1.5	Contribution	2
1.6	Organization of the report	3
2	Literature Survey	4
2.1	Introduction	4
2.1.1	Football match prediction using player attributes	4
2.1.2	An Improved Prediction System for Football a Match Result . .	4
2.2	Problem Definition	5
3	Design Implementation	6
3.1	Proposed System	6
3.2	Requirement Gathering and Analysis	6
3.3	Hardware Requirement	7
3.4	Software Requirement	7
3.5	Data Flow Diagram	7
3.5.1	Data Flow Diagram	7
3.6	Cost Estimation	7
3.7	Feasibility Study	8
4	Results and Discussion	9
4.1	Code	9
4.2	Additional details of the Project	9
4.3	Screen Shots	10
4.4	Testing Results	12
5	Conclusion	13
5.1	Summary	13
5.2	Future Scope	13

List of Figures

3.1	Data Flow Diagram	7
4.1	results of various models	12

Chapter 1

Introduction

1.1 Introduction

The English Premier League is one of the most popular and competitive football leagues in the world, attracting millions of fans and generating significant economic activity. Predicting the outcome of Premier League matches has long been a topic of interest for fans, analysts, and bookmakers. The global sports betting market is expected to reach over 155 billion dollars by 2024, with online betting accounting for a growing share of the market. In the UK, online betting accounts for over 70 percent of total betting activity, with football being one of the most popular sports for betting. The popularity of sports betting has led to an increase in the number of betting companies, with many offering highly sophisticated betting tools and services. This project aims to develop a machine learning model that can accurately predict the outcome of English Premier League matches. By analyzing data from past matches and incorporating current team and player information, the model will provide actionable insights that can help stakeholders make informed decisions. The model will also continuously learn and adapt to changing circumstances, providing up-to-date predictions in real-time.

1.2 Aim and Objectives

The project aims to develop a machine learning model that can accurately predict the outcome of English Premier League matches.

1. Predict the outcome of English Premier League matches with high accuracy.
2. Identify key factors that influence the outcome of matches, such as team statistics.
3. To improve the model's performance and increase accuracy over time.
4. Develop a user-friendly interface to enable easy access and interpretation of the model's predictions and insights.
5. Continuously update the model to incorporate recent match data and current team and player information.

1.3 Motivation for the Work

The motivation for this project is to create a useful tool for football fans and analysts who want to predict the outcomes of football matches. The project will also contribute to the growing field of machine learning and sports analytics, which has the potential to revolutionize how we analyze and understand sports.

1.4 Scope of Project

The scope of this project includes collecting and preprocessing football match data, developing and testing various machine learning models, evaluating their performance, and developing a user-friendly interface for the predictor. The project will later also focus on predicting the outcomes of matches in major football leagues worldwide. The predictor will take into account various factors that can influence match outcomes, such as team form, player injuries, and head-to-head records.

1.5 Contribution

Overall, an English Premier League match predictor can make significant contributions to the world by enhancing the viewing experience for Football fans, facilitating informed decision-making for sports enthusiasts, advancing the use of data analytics and machine learning in sports, and supporting the growth and competitiveness of the sports industry.

There are a number of ways in which our match predictor can contribute to the world.

1)Enhancing the viewing experience for Football fans: By providing accurate predictions of match outcomes and generating insights into key factors that influence results, the match predictor can help Football fans better understand and appreciate the game.

2)Facilitating informed decision-making for sports enthusiasts: The match predictor can assist sports enthusiasts in making informed betting decisions based on data-driven insights, thereby minimizing the risk of losing money due to uninformed decisions

3)Advancing the use of data analytics and machine learning in sports: By leveraging the power of data analytics and machine learning, the match predictor can advance the use of these technologies in the sports industry, potentially leading to the development of more sophisticated and accurate predictive models.

4)Supporting the growth and competitiveness of the sports industry: By providing accurate predictions of match outcomes, the match predictor can help sports teams and organizations better understand their performance and identify areas for improvement, potentially leading to greater competitiveness and growth in the industry.

1.6 Organization of the report

This report is organized into the following sections:

1. Introduction
2. Research Methodology
3. Executive Summary
4. Description of the industry

1. Introduction

1. Brief overview of the English Premier League
2. Explanation of the purpose and scope of the report
3. Description of the match predictor and its potential benefits

2. Research Methodology

1. Description of the data sources used to train the predictor
2. Explanation of the machine learning algorithms used
3. Overview of the feature selection process
4. Discussion of the model evaluation metrics

3. Executive Summary

1. Overview of the predictor's accuracy and performance
2. Explanation of the key findings and insights generated by the predictor
3. Description of the potential applications and benefits of the predictor

4. Description of the industry

1. Overview of the Football industry, including its history and current state
2. Discussion of the key trends and challenges facing the industry
3. Overview of the role of data analysis and machine learning in the industry

Chapter 2

Literature Survey

2.1 Introduction

A literature survey is a comprehensive review of existing literature, research, and other relevant sources on a particular topic. For this English Premier League match predictor project, the literature survey involves identifying and examining existing literature and research related to the prediction of EPL matches. In this literature survey, we are considering a range of sources, including academic articles, books, conference proceedings, and online resources. We also consider sources that address related topics, such as sports analytics, data mining, and machine learning.

2.1.1 Football match prediction using player attributes

Home advantage effect on team's performance

Home advantage is “the consistent finding that the home team wins over 50 percent of the games played under balanced home and away schedule” . Categorically, travel and crowd factors could also account for the home advantage effect on team's performance. Experimental study by reported the evidence of home team favouritism in the English Premier League and German Bundesliga. More so, and drew the same conclusion based on their respective studies: that it makes sense to model home and away performance separately.

2.1.2 An Improved Prediction System for Football a Match Result

Effect of Injuries

The effect of injuries of key player on team performance reviewed the effect of injuries of individual players on the performance of the entire team. In their analysis, they measured the team quality as a sum of players' quality scores from a sample of five seasons of English Premier League. This research showed that the injuries of key players plays significant role on team's performance. However, argued that such information is subjective in nature and should not be considered in statistical forecasting.

Effect of external cup on league performance

The explanation to the effect of external cup can be described by the financial incentives cups offer. , propounded that teams are more interested to perform well in

league with direct revenue than that without direct financial incentives. , claimed that cup participation would imply an increase in league performance, consequently a cup exit would imply decrease in league performance.

2.2 Problem Definition

In the world of sports, predicting the outcome of a match is a challenging task that requires a deep understanding of the various factors that influence the result. The English Premier League is one of the most competitive and widely followed football leagues in the world, and accurately predicting the outcome of its matches can provide valuable insights for fans, bookmakers, and analysts. The goal of this project is to develop a machine learning model that can accurately predict the outcome of English Premier League matches, taking into account factors such as team statistics, player performance, and injury information. The model should be able to provide actionable insights that can help improve its performance over time and help stakeholders make informed decisions.

Chapter 3

Design Implementation

3.1 Proposed System

Data collection: Collect data on past EPL matches, including match statistics such as goals scored, shots on target, possession, and other relevant data. You can use APIs such as the Football-Data API or web scraping techniques to collect this data.

Data preprocessing: Clean and preprocess the collected data. This step includes handling missing values, encoding categorical variables, and feature engineering.

Model selection: Select an appropriate machine learning model for predicting EPL matches. You can use algorithms such as logistic regression, decision trees, random forests, or neural networks.

Model training: Train the selected model on the preprocessed data. Use cross-validation techniques to evaluate the model's performance.

Model testing: Test the trained model on a separate test dataset to evaluate its performance.

Hyperparameter tuning: Tune the model's hyperparameters to improve its performance.

Deployment: Deploy the model to make predictions on new EPL matches.

User interface: Develop a user interface for the EPL match predictor system. This interface should allow users to input the two teams playing in a match, and the system should provide a predicted outcome.

3.2 Requirement Gathering and Analysis

Building an English Premier League match predictor would require expertise in data analysis, machine learning, and web development. The project would involve several steps, including data retrieval, preprocessing, analysis, model training, and evaluation. By utilizing the right tools and technologies, developers can create a user-friendly predictor that provides accurate predictions for Premier League matches.

3.3 Hardware Requirement

Windows 8/10 (32/64 bit) or higher

Minimum 4 GB RAM and higher

10 GB available space on the hard disk

Active internet connection minimum speed 512 kbps and above.

3.4 Software Requirement

A Python notebook IDE.

Internet Browser e.g. Chrome, Firefox, Microsoft Edge etc.

3.5 Data Flow Diagram

3.5.1 Data Flow Diagram

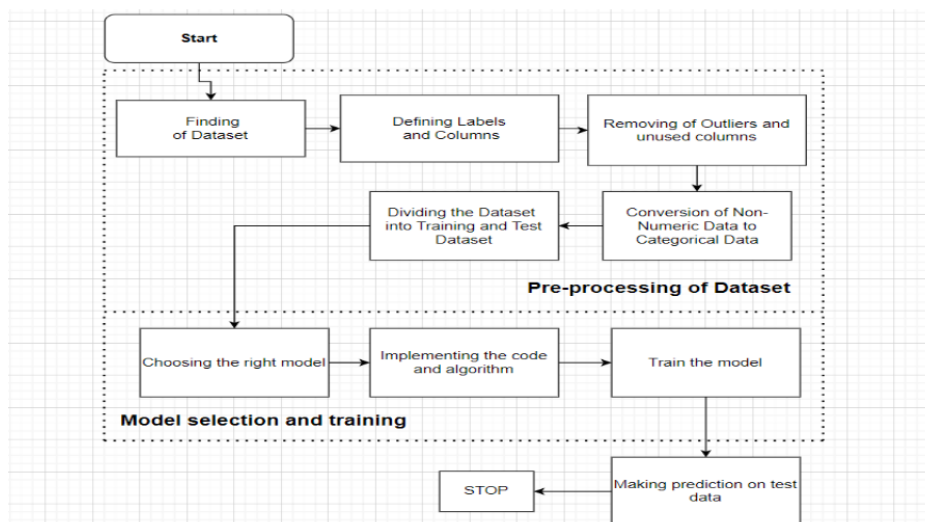


Figure 3.1: Data Flow Diagram

3.6 Cost Estimation

There are not many variables to consider when estimating the cost of a match predictor system. Some variables would be extra ram space to work with large datasets efficiently. Also the development team must consider the cost of developing the predictor, including the cost of data acquisition, computational resources, software licenses, and personnel costs.

3.7 Feasibility Study

Technical feasibility:

Availability of data: The predictor will require access to reliable and comprehensive data on team and player performance, weather conditions, and other relevant factors. This data must be available in a structured format that can be easily processed and analyzed.

Machine learning expertise: The development team must have expertise in machine learning algorithms, data analysis, and statistical modeling to design and train an accurate prediction model.

Computational resources: The predictor will require significant computational resources for data preprocessing, analysis, and model training. The team must have access to sufficient computing power to perform these tasks efficiently.

Economic feasibility:

Cost of development: The development team must consider the cost of developing the predictor, including the cost of data acquisition, computational resources, software licenses, and personnel costs.

Revenue potential: The predictor may generate revenue through advertising, premium features, or subscription services. The team must evaluate the revenue potential of the predictor and determine whether the potential revenue outweighs the development costs.

Operational feasibility:

User adoption: The predictor must be user-friendly and provide accurate predictions to gain user adoption. The development team must consider user feedback and iterate on the predictor to improve accuracy and user satisfaction.

Maintenance and updates: The predictor will require ongoing maintenance and updates to stay up-to-date with the latest data and performance trends. The development team must plan for ongoing maintenance and support to ensure the predictor remains accurate and reliable.

Overall, an English Premier League match predictor in Python is technically feasible given the availability of data and machine learning expertise. However, the economic feasibility will depend on the revenue potential and development costs, while operational feasibility will depend on user adoption and ongoing maintenance and updates.

Chapter 4

Results and Discussion

4.1 Code

App.js

```
from sklearn.neighbors import KNeighborsRegressor
knnr = KNeighborsRegressor(n_neighbors=5, weights='uniform')
knnr.fit(Xhome_train, yhome_train)

y_pred = knnr.predict(Xhome_test)

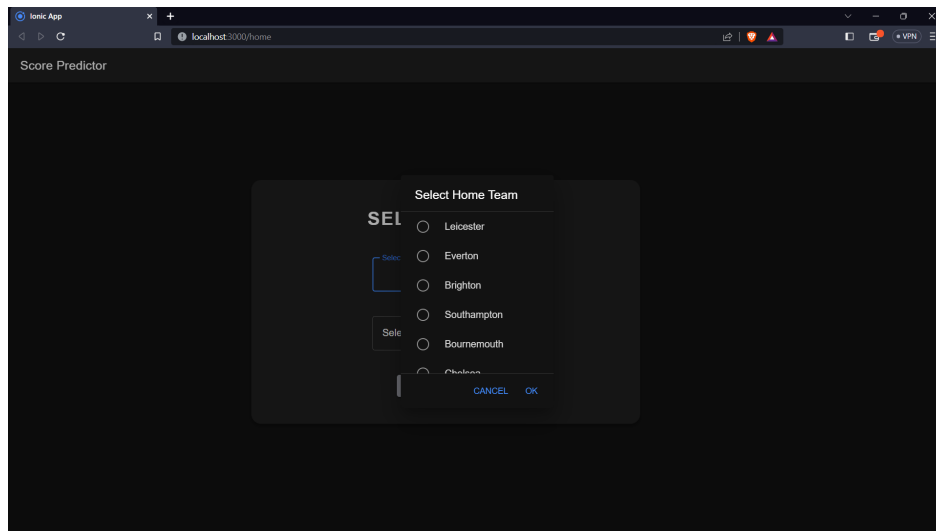
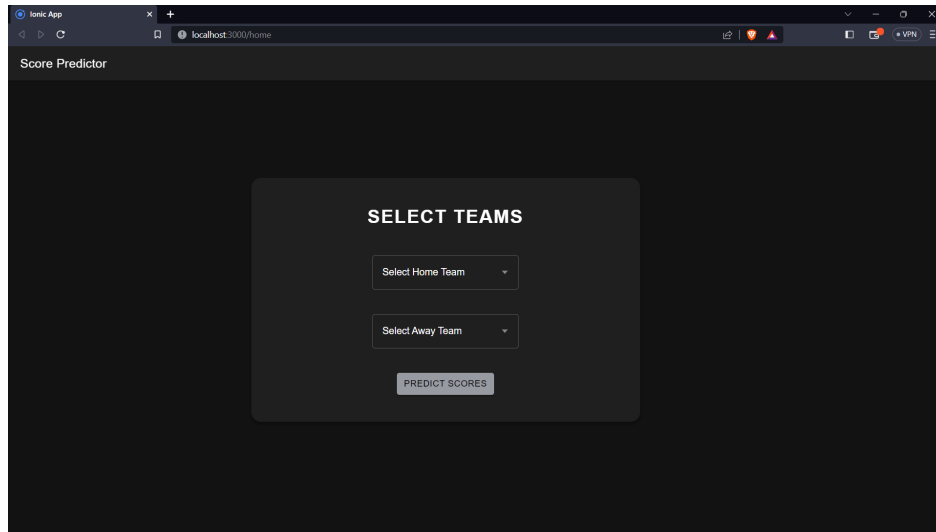
print('Mean Squared Error:', mean_squared_error(yhome_test, y_pred))
print('r2 score: ', r2_score(yhome_test, y_pred))
```

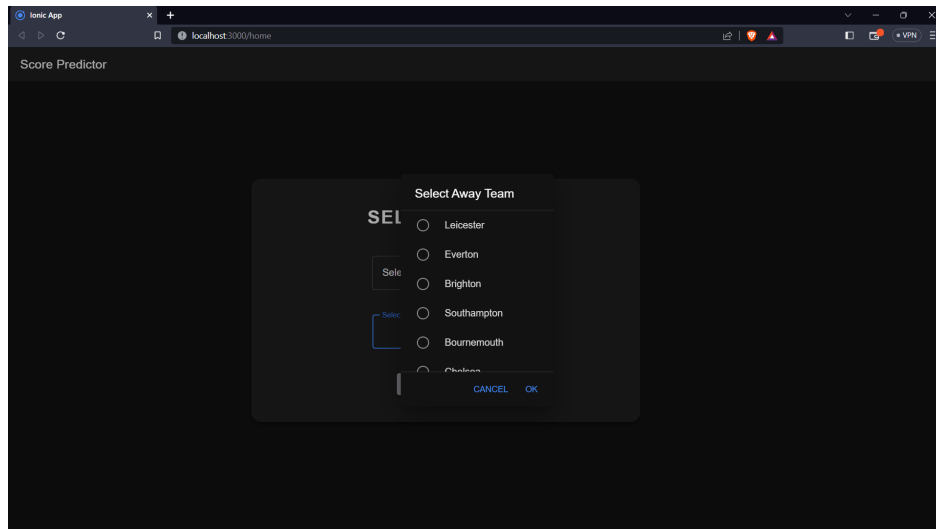
Whole code can be accessed using this link.

4.2 Additional details of the Project

Feature selection is an important step in building a prediction model as it helps to identify the most relevant factors that can influence the outcome of the match. Some relevant factors may include team form, home or away advantage, player injuries or suspensions, weather conditions, and historical head-to-head records. Exploratory data analysis techniques such as correlation analysis, feature importance analysis, and data visualization can be used to identify and select the most relevant features.

4.3 Screen Shots





4.4 Testing Results

Model	R2_Score	Prediction (%)
SVR	0.9974	27.86
Lasso	0.9937	64.18
Random Forest	0.9366	55.72
KNN	0.9901	95.52
Elastic Net	0.9927	64.18
Ridge	0.9999	54.73

Figure 4.1: results of various models

Chapter 5

Conclusion

5.1 Summary

The English Premier League Match Predictor is a data-driven project that uses machine learning algorithms to predict the outcomes of football matches in the English Premier League. The project involves collecting and analyzing historical data on teams and players, selecting relevant features that impact match outcomes, and training a predictive model. The model is then used to generate accurate predictions of future match outcomes, which can be used by Football fans to enhance their viewing experience and by sports enthusiasts to make informed betting decisions. The project also contributes to the advancement of data analysis and machine learning in the sports industry, potentially leading to greater competitiveness and growth.

5.2 Future Scope

Improving predictability using neural networks.

Predicting various different attributes about the match such as injuries,fouls etc.

Predicting using more attributes such as external leagues,transfers etc.

References

- 1 Norbert Danisik, Peter Lacko, Michal Farkas Faculty of Informatics and Information Technologies Slovak University of Technology Bratislava, Slovakia “Football match prediction using player attributes”,2016
- 2 “An Improved Prediction System for Football a Match Result” Igiri, Chinwe Peace ; Nwachukwu, Enoch Okechukwu, 12 Dec 2014,