

Football Match Prediction using Machine Learning

**Submitted in partial fulfillment of the requirement for the award of
Degree of Bachelor of Technology in Computer Engineering Discipline**

Submitted To



**SVKM's NMIMS,
Mukesh Patel School of Technology Management & Engineering,
Shirpur Campus (M.H.)**

Submitted By:

**Pranav Kolhe – 70021119035
Abhraneel Das – 70021119011
Dhairya Desai – 70021119012**

Under The Supervision Of:

**Dr. Upendra Verma
(Asst. Professor, Computer Engineering)**

**DEPARTMENT OF COMPUTER ENGINEERING
Mukesh Patel School of Technology Management & Engineering
SESSION: 2022-23**

CERTIFICATE

This is to certify that the work embodies in this Project entitled “**Football Match Prediction using Machine Learning**” being submitted by

Pranav Kolhe – 70021119035

Abhraneel Das – 70021119011

Dhairya Desai – 70021119012

for partial fulfillment of the requirement for the award of “**Bachelor of Technology in Computer Engineering**” discipline to “SVKM’s NMIMS, Mumbai (M.H.)” during the academic year 2022-23 is a record of bonafide piece of work, carried out by them under my supervision and guidance in the “**Department of Computer Engineering**”, MPSTME, Shirpur (M.H.).

APPROVED & SUPERVISED BY:

Dr. Upendra Verma
(Asst. Professor, Computer Engineering)

FORWARDED BY

(Dr. N.S. Choubey)
H.O.D., Computer
MPSTME, Shirpur Campus

(Dr. Venkatadri M.)
Asso. Dean, Engg. Programs,
MPSTME, Shirpur Campus

DEPARTMENT OF COMPUTER ENGINEERING
Mukesh Patel School of Technology Management & Engineering

CERTIFICATE OF APPROVAL

The Project entitled “**Football Match Prediction using Machine Learning**” being submitted by

Pranav Kolhe – 70021119035

Abhraneel Das – 70021119011

Dhairya Desai – 70021119012

has been examined by us and is hereby approved for the award of degree “**Bachelor of Technology in Computer Engineering Discipline**”, for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein, but approve the project only for the purpose for which it has been submitted.

(Internal Examiner)

(External Examiner)

Date:

Date:

**DEPARTMENT OF COMPUTER ENGINEERING
Mukesh Patel School of Technology Management & Engineering**

DECLARATION

We,

Pranav Kolhe – 70021119035

Abhraneel Das – 70021119011

Dhairya Desai – 70021119012

The students of **Bachelor of Technology in Computer Engineering discipline, Session: 2022-23, MPSTME, Shirpur Campus**, hereby declare that the work presented in this Project entitled “**Bachelor of Technology in Computer Engineering Discipline**” is the outcome of our work, is bonafide and correct to the best of our knowledge and this work has been carried out taking care of Engineering Ethics. The work presented does not infringe any patented work and has not been submitted to any other university or anywhere else for the award of any degree or any professional diploma. [14pt]

Pranav Kolhe – 70021119035

Abhraneel Das – 70021119011

Dhairya Desai – 70021119012

Date:

DEPARTMENT OF COMPUTER ENGINEERING
Mukesh Patel School of Technology Management & Engineering

ACKNOWLEDGEMENT

After the completion of this Major Project work, words are not enough to express our feelings about all those who helped us to reach my goal; feeling above this is our indebtedness to The Almighty for providing us this moment in life.

It's a great pleasure and moment of immense satisfaction for us to express our profound gratitude to **Dr. Upendra Verma, Asst. Professor**, Computer Engineering Department, MPSTME, Shirpur, whose constant encouragement enabled us to work enthusiastically. Their perpetual motivation, patience and excellent expertise in discussion during progress of the project work have benefited us to an extent, which is beyond expression. Their depth and breadth of knowledge of Computer Engineering field made us realize that theoretical knowledge always helps to develop efficient operational software, which is a blend of all core subjects of the field. We are highly indebted to them for their invaluable guidance and ever-ready support in the successful completion of this project in time. Working under their guidance has been a fruitful and unforgettable experience.

We express our sincere thanks and gratitude to Dr. N.S. Choubey, Head of Department, Computer Engineering Department, MPSTME, Shirpur, for providing necessary infrastructure and help to complete the project work successfully.

We also extend our deepest gratitude to Dr. Akshay Malhotra, Director, SVKM'S NMIMS, Shirpur Campus and Dr. Venkatadri M., Associate Dean, Engineering Programs, SVKM'S NMIMS, Shirpur Campus for providing all the necessary facilities and true encouraging environment to bring out the best of our endeavors.

We sincerely wish to express our grateful thanks to all members of the staff of computer engineering department and all those who have embedded us with technical knowledge of computer technology during various stages of B.Tech. Computer Engineering.

We would like to acknowledge all of our friends, who have contributed directly or indirectly in this Major Project work.

The successful completion of a Major Project is generally not an individual effort. It is an outcome of the cumulative effort of a number of persons, each having their own importance to the objective. This section is a vote of thanks and gratitude towards all those persons who have directly or indirectly contributed in their own special way towards the completion of this project.

Pranav Kolhe – 70021119035

Abhraneel Das – 70021119011

Dhairya Desai – 70021119012

ABSTRACT

One of the most popular sports in the world is Football. Millions of people watch Football every day. Big clubs like Manchester City, Manchester United, Barcelona, Real Madrid, Juventus, Inter Milan, Bayern Munich, etc. have a huge fan following. With such a huge fan following comes huge expectations. Every fan wants its team to win.

This makes the Predictions of Football matches even more popular, be it the manager, staff, pundits or even the fans themselves. Football is one of the most far reaching and famous game, along these lines anticipating the consequences of a football match represents a fascinating test.

Forecast can likewise help administrators and clubs in settling on the best choices to win associations and competitions. In view of the huge financial stakes engaged with wagering, there is an interest for high prescient exactness in this industry.

In recent times, a lot of Research is been done in Football Predictions. Many Machine Learning and Deep Learning Algorithms are used to predict which team is going to win the match.

It is been estimated that Football is played over 200 countries and 1.4 billion Football fans are supporting their teams globally. European Football alone exceeds 28 billion EUROS market scale. This forces the football team management to select proper strategy for each match. This is done with a team of analysts analyzing the data of their players, as well as their opponents.

Manual evaluation tends to generate many errors and take more time. In the current research the statistical model is proposed to predict the stats of the football player based on previous session data by considering various aspects of the game.

Our objective is to predict the outcomes of a Football match.

We plan to use and compare multiple Machine Learning models, to find out which of the above models used can accurately predict the result of the match. We are going to use various parameters from the dataset to improve the accuracy of the model.

TABLE OF CONTENTS

S. NO.	CHAPTER	PAGE
1	INTRODUCTION	1-2
	1.1 Purpose	1
	1.2 Scope	1
	1.3 Overview	2
2	LITERATURE SURVEY	3 – 9
3	PROBLEM DEFINITION & PROPOSED SOLUTION	10
	3.1 Problem Statement	
	3.2 Proposed Solution	
4	DESIGN	11 – 16
	4.1 Architectural Diagram	11
	4.2 Data Flow Diagram	12-13
	4.3 Use Case Diagram	13
	4.4 Activity Diagram	14
	4.5 Class Diagram	14
	4.6 Sequence Diagram	15
	4.7 Component Diagram	15
	4.8 Collaboration Diagram	16
	4.9 State Chart Diagram	16
5	Result Analysis	17 - 21
	5.1 Home Win Analysis	18
	5.2 Away Win Analysis	19
	5.3 Draw Analysis	20
	5.4 Model Accuracy	21
6	Testing	22 – 24
	6.1 Unit Testing	22
7	Conclusion and Future Work	25
	7.1 Conclusion	
	7.2 Future Work	
	References	26 - 30

LIST OF FIGURES

Sr. No.	Figure No.	Figures	Page
1	4.1	Architecture Diagram	11
2	4.2.1	Data Flow Diagram Level - 0	12
3	4.2.2	Data Flow Diagram Level - 1	12
4	4.2.3	Data Flow Diagram Level - 2	13
5	4.3	Use Case Diagram	13
6	4.4	Activity Diagram	14
7	4.5	Class Diagram	14
8	4.6	Sequence Diagram	15
9	4.7	Component Diagram	15
10	4.8	Collaboration Diagram	16
11	4.9	State Chart Diagram	16
12	5.1	Impact of Home Wins Graph	17
13	5.2	Home Win Analysis Graph	18
14	5.3	Away Win Analysis Graph	19
15	5.4	Draw Analysis Graph	20
16	5.5	Model Accuracy Graph	21
17	6.1	Test Case - 1	22
18	6.2	Test Case - 2	23

CHAPTER 1

INTRODUCTION

1.1 PURPOSE

The purpose of our project is to successfully create a computer model that can successfully predict whether a football team would win, lose, or draw a match. By observing patterns, asking questions, and building a model, we will have a chance to push state-of-the-art technology farther.

The document is intended for sports analysts and researchers. It is also intended for football club managers and staff members. This project is a prototype model for football match prediction. It will be very useful for these individuals and will assist them by saving their time by predicting the match results and adjusting their strategy accordingly.

1.2 SCOPE

Preparing a perfect strategy for their own team in an important and also a tedious task in sports industry. This also involves predicting and analyzing every possible scenario of the future matches.

Many researchers have attempted to predict football match outcome using the in-game match statistics. There is so much data available other than that.

Hence, creating a computer model that can predict the outcome of a particular football match would be of a great help. Predicting the outcomes of a match can help the analysts of a football team simulate a few strategies and pick the best ones for each match.

1.3 OVERVIEW

The aim of this document is to provide an overall description of the project Football Match Prediction using Machine Learning. Starting off with the Introduction of the project, the reader understands the purpose and the scope of the project. Also, if needed, the reader can check the references for more information on the topic. Short-forms that might not be known to the audience are also mentioned.

Moving on to the next part, a general description of the functionality, context, and design of the project is provided. The reader can go through it to understand the procedure that is to be followed throughout the process.

First the system diagram is provided, that contains Architectural Design and Decomposition description. The document shows Object-Oriented Approach towards the project and Human Interface Design. The Object-Oriented Approach has 8 diagrams: Use Case Diagram, Class Diagram, Activity Diagram, Sequence Diagram, Collaboration Diagram, State Chart Diagram, Component Diagram and Deployment Diagram.

Further, the results and testing are shown to provide understanding of the coding section of the project. Finally, the document is ended with Conclusion and References.

LITERATURE SURVEY

2.1 RESEARCH PAPERS

We have referred several research papers which are as follows:

(1) Yoel F. Alfredo, Sani M. Isa, **Football Match Prediction with Tree Based Model Classification**, 2019

Here, Random Forest, XGBoost and C 5.0 algorithms were used and we found out that Random Forest algorithm generated the best accuracy with 68,55% while the C5.0 algorithm had the lowest accuracy at 64,87% and Extreme Gradient Boosting algorithm produced accuracy of 67,89%. The limitation of this paper was that a limited dataset was used and a better feature selection method would have resulted in a better result.

(2) Syasya Nadhilah Maozad, Siti Noor Asyikin Mohd Razali, Aida Mustapha, Aziz Nanthaamornphong, Mohd Helmy Abdul Wahab, Nazim Razali, **Comparative Analysis for Predicting Football Match Outcomes based on Poisson Models**, 2022

Here, Dixon-Coles model, Rue-Salvesen model, and Poisson Regression model were used and the experiments showed that the best prediction model is the Dixon-Coles Poisson model with the highest log-likelihood of -15023.29 and the lowest AIC value of 32056.57. It has an accuracy of 82%.

(3) Hang Li, **Analysis on the construction of sports match prediction model using neural network**, 2020

Here, back-propagation neural network (BPNN), multiple linear regression (MLR) model and grey degree prediction model were used, RMSE in training and prediction stage is 0.22 and 0.34, respectively, MAPE is 7.60% and 9.21%, respectively, and MSE is 0.89 and 0.85, respectively. Thus the improved adaptive BPNN has higher fitting accuracy and better performance.

(4) Jan Kozaka,Szymon Głowaniaa, **Heterogeneous ensembles of classifiers in predicting Bundesliga football results**, 2021

Here, bagging, random forest, ADA Boost were used and it was observed that best rank of 1.4286 was obtained for proposed method that is the version with selections. But the major limitation in this paper was that there was a serious drawback with the prediction of Draw Result.

(5) Che Mohamad Firdaus, Che Mohd Rosli, **A Comparative Study of Data Mining Techniques on Football Match Prediction**, 2018

Here, Decision Trees, Neural Networks, Bayesian networks, K-nearest neighbours were used and we can observe that Decision Trees have the highest average accuracy by 99.56% follow by Neural Networks and k-Nearest Neighbours technique by 96.83% and 77.54% while Bayesian Networks have the worst average accuracy by 76.41%. The limitation mentioned in this paper was that limited attributes and parameters were selected. More parameters would have given more accurate results.

(6) Yash Ajgaonkar,Prof. Anagha Patil,Kunal Bhoyar,Jenil Shah, **Prediction of Winning Team using Machine Learning**, 2020

Here, SVM, Random Forest, Naive Bayes were used and the accuracy obtained for Support Vector Machine is 67% which is comparatively greater than Random Forest and Naïve Bayes obtained as 60% and 56% respectively. Here, the limitation was that the model is not sensitive to sentiment analysis that is prediction is done on purely statistical basis.

(7) Mrigank Vashist, Vasudha Bahl, Nidhi Sengar, Amita Goel, **Machine Learning for Football Matches and Tournaments**, 2022

Here, Ensemble learning was used and the model had an accuracy of 94.8%. Limitation in this paper was that the model was tested on data from a singular league so it is dicey if it will work well with other leagues as well.

(8) Hengzhi Chen, **Neural Network Algorithm in Predicting Football Match Outcome Based on Player Ability Index**, 2019

Here, Convolutional neural network (CNN), random forest (RF) and support vector machine(SVM) were used and the accuracy of these three methods is all between 54% and 58%. Limitation in this paper is that all the three models seldom predict draw. Also, bifurcation can be done on the basis of the model's ability to predict win or loss that is one model is better at predicting win and vice versa.

(9) Riccardo Ievoli, Lucio Palazzo , Giancarlo Ragozini, **On the use of passing network indicators to predict football outcomes**, 2021

Here, Binomial Logistic Regression (BLR) was used and it presents best performance in terms of accuracy 76% and AUC 81%. Limitation of this paper is that the model did not focus on adding various spatial information.

(10) H R Azhari, **Predicting Final Result of Football Match Using Poisson Regression Model**, 2018

Here, Poisson Regression was used and had an accuracy of 61%.

(11) Chananyu Pipatchatchawal, Suphakant Phimoltares, **Predicting Football Match Result Using Fusion-based Classification Models**, 2021

Here, fusion based classification, hierarchical model, ensemble model was used and we can see that the both proposed models yielded the accurate rates at 56.5332% and 56.8002%. The limitation mentioned in this paper was that limited attributes and parameters were selected. More parameters would have given more accurate results.

(12) Shuo Guan, Xiaochen Wang, **Optimization analysis of football match prediction model based on neural network**, 2021

Here, Gray fuzzy prediction was used and the prediction accuracy of the two combined models is above 70%.

(13) Darwin Prasetyo , Dr. Harlili, **Predicting Football Match Results with Logistic Regression**, 2021

Here, Logistic Regression was used and the prediction accuracy of built model is 69.5%. Limitation here is that the coefficients estimation on the constant tend to have bigger numbers.

(14) Yan Watequlis Syaifudin, Pradini Puspitaningayu, **Predicting Winner of Football Match Using Analytical Hierarchy Process: An Analysis Based on Previous Matches Data**, 2021

Here, Analytical Hierarchy Process was used and the effectiveness of the applied method is 86.67%. The limitation mentioned in this paper was that limited attributes and parameters were selected. More parameters would have given more accurate results.

(15) Jassim Almulla, Tanvir Alam, **Machine Learning Models Reveal Key Performance Metrics of Football Players to Win Matches in Qatar Stars League**, 2020

Here, KNN, Logistic Regression, Random Forest, XG Boost Classifier, Decision Tree, MLP Classifier, SVM were used and out of all the models, Logistic Regression achieved the highest accuracy of 80.1%. Limitation is that Limited Data was used in this model. A larger dataset would have improved the model accuracy. Also, Home/Away factor, grass and ball types, etc were not included.

(16) Azrel Aiman Azeman, Aida Mustapha, Nazim Razali, Aziz Nanthamornphong, Mohd Helmy Abd Wahab, **Prediction of Football Matches Results: Decision Forest against Neural Network**, 2021

Here, Multiclass Neural Network, multiclass Decision Forest were used and we found out that Decision Forest had highest accuracy of 88.95%. Here, Random Forest model would have been more accurate and would have produced a better result than the models used in this paper on the dataset.

(17) Dongdong Wang, **Clustering and Evolutionary System Analysis of Data Mining Algorithms in the field of Football**, 2022

Here, Data clustering through local metric learning, FKM clustering method were used and found out that out of a testing pool of 6 sportsmen the winning rate percentages were ranging from 70-80%. The results would have been better if the economic benefits were to be analysed too.

(18) Syasya Nadhilah Maozad, Siti Noor, Asyikin Mohd Razali, Aida Mustapha Aziz, Nanthaamornphong, Mohd Helmy Abdul Wahab Nazim Razali **Comparative Analysis for Predicting Football Match Outcomes based on Poisson Models**, 2022

In this one, Dixon-Coles model, Rue-Salvesen model, and Poisson Regression model was seen to be used and they chose DixonColes model because the value of log-likelihood is the highest meanwhile the value of AIC is the lowest. It has an accuracy of 82%. The models may differ in result if were used to predict expected goal score.

(19) S Manish, Vandana Bhagat, RM Pramila, **Prediction of Football Players Performance using Machine Learning and Deep Learning Algorithms**, 2021

Here, Performance Prediction, Deep Learning Algorithm, Machine Learning algorithm, Linear Regression, Neural networks were used and we found out that the mean squared error for the defence model is 21.3 although the R squared value is 0.90. The R squared values from the XGBoost algorithm is 0.90. The multiple linear regression results are better than other regression model for the experimental data.

(20) S.K. Nivetha, M. Geetha, R.C. Suganthe, R Manoj Prabhakaran, S Madhuvanan, A Mohamed Sameer, **A Deep Learning Framework for Football Match Prediction**, 2022

Here, LSTM (Long Short-Term Memory), RNN (Recurrent Neural Networks), Deep Learning models were used, in which the LSTM model gives the highest accuracy of 86%.

(21) Md. Ashiqur Rahman, **A deep learning framework for football match prediction**, 2020

Here, LSTM, ANN and DNN were used and the algorithms manages to get an accuracy of 63.3%. More accurate data was needed. The model could not predict the Quarter-Finals, Semi-Finals, Final matches accurately.

(22) N. Danisik, P. Lacko and M. Farkas, **Football Match Prediction Using Players Attributes**, 2018

Here, LSTM, regression models and RNN were used and out of all them LSTM model provides the highest accuracy of 52.48%. In-depth analysis like specific tactics or different relationships in time aspect of the football games played could have improved the accuracy of the model.

(23) Qiyun Zhang, Xuyun Zhang, Hongsheng Hu, Caizhong Li, Yinping Lin, Rui Ma, **Sports match prediction model for training and exercise using attention-based LSTM network**, 2022

Here, along with DNN, AS-LSTM was used and it provided an accuracy of 80%. A more accurate athlete-oriented dataset would have improved the model accuracy.

(24) Nilay Zaveri, Shubham Tiwari, Pramila Shinde, Utkarsh Shah, Lalit Kumar Teli, **Prediction of Football Match Score and Decision-Making Process**, 2018

Here, Logistic Regression, Linear SVM, ANN, Naïve Bayes Classifier and Random Forest were used and logistic Regression provides the highest accuracy of 71.63%. The drawback of the system is that it uses a very small dataset (96 matches) i.e., the data of 1 complete UCL season and the testing is done on the same.

(25) Laura Hervert-Escobar, Neil Hernandez-Gress, Timothy I. Matis, **Bayesian Based Approach Learning for Outcome Prediction of Soccer Matches**, 2018

Here, Basic Bayesain mdel was used and the ranked Probability score ranges from 33-40%. Knowledge-based system should be used.

(26) Muhammad Haleq Azhar Abdul Rahman, Aida Mustapha, Rahmat Fauzi, Nazim Razali, **Bayesian Approach to Classification of Football Match Outcome**, 2018

Here, Naïve Bayes, Tree Augmented Naïve Bayes and General Bayesian Network (K2) were used and we saw that the Tree Augmented Naïve Bayes gets the highest accuracy of 90%. Other Bayesian learning algorithm in general Bayesian Network should also be tried.

(27) Tetyana Korotyeyeva, Ruslan Tushnyskyy, Volodymyr Kulyk, **Applying Neural Networks to Football Matches Results Forecasting**, 2018

Here, Neural Network, K-median, Naïve Bayes were used and we found out that 75% accuracy of prediction was obtained. The modified algorithm such as using a greater number of hidden layers can improve the accuracy.

(28) Ekansh Tiwari, Prasanjit Sardar, Sarika Jain, **Football Match Result Prediction Using Neural Networks and Deep Learning**, 2020

Here, RNN and LSTM were used and LSTM form of RNNs produced an accuracy of 80.75%. This model should be tried on different data, parameters and inputs.

(29) Laura Hervert-Escobar, Neil Hernandez-Gress and Timothy I. Matis, **Prediction Learning Model for soccer matches outcomes**, 2018

Here, Mixure of Dynamic ratings and Hybrid Bayes Algorithms were used and the ranked Probability score ranges from 27%-33%. More data should be considered to get accurate prediction of a draw match.

(30) Fatima Rodrigues, Angelo Pinto, **Prediction of football match results with Machine Learning**, 2022

Here, SVM, Random Forest, XG Boost, RNA was used and we found out that Random Forest was the best model with a prediction accuracy of 65.26%. The forecast model should be integrated in a decision support system based on the probability of occurrence of the forecast model results.

CHAPTER 3

PROBLEM DEFINITION & PROPOSED SOLUTION

3.1 PROBLEM STATEMENT

Machine Learning based Football Match Prediction is crucial for Football Team analysts and Strategists to make the right call with respect to how their team would perform.

The main objective of the project is to create an ML System that can help the team Analysts make accurate decisions.

3.2 PROPOSED SOLUTION

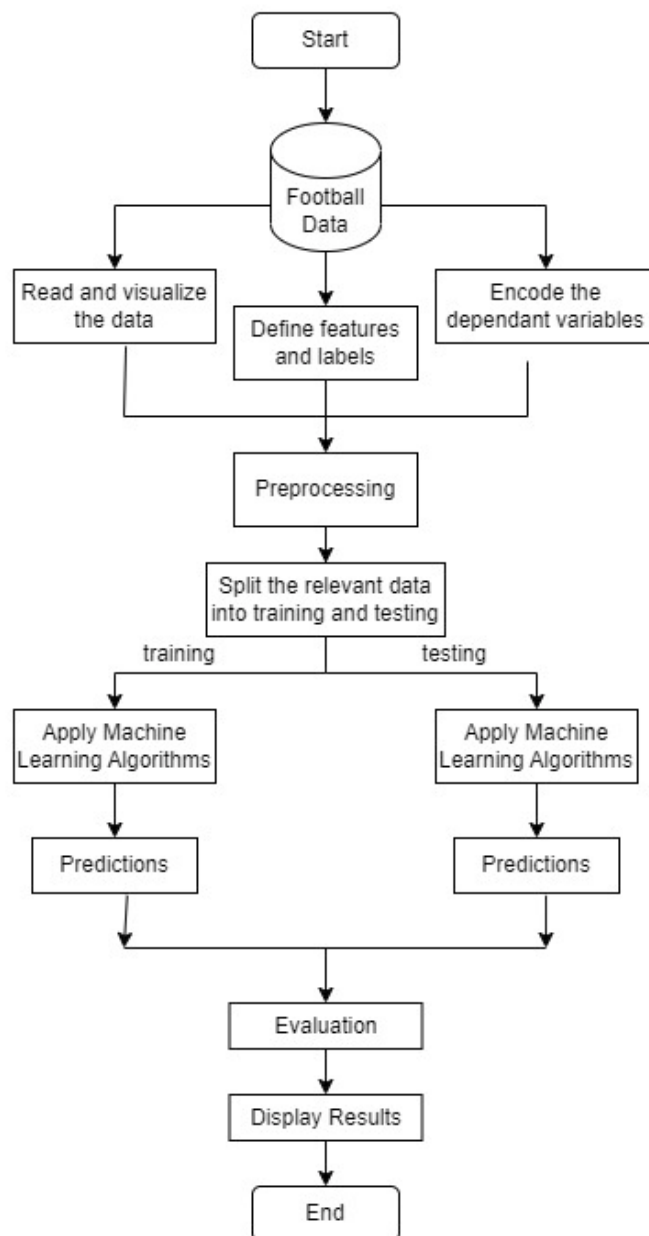
Using multiple Machine Learning Algorithms, the model with the highest accuracy for win-loss-draw results would be considered in our project as the best predictive model for Football Match Predictions.

And then that selected predictive model will then be used to create a graphical presentation for the supposed football team analysts for predicting the nearest possible solution for optimum victory against the opposition.

CHAPTER 4

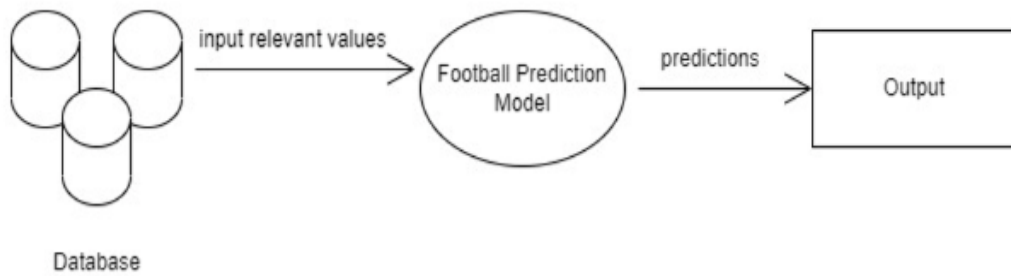
DESIGN

4.1 ARCHITECTURAL DIAGRAM

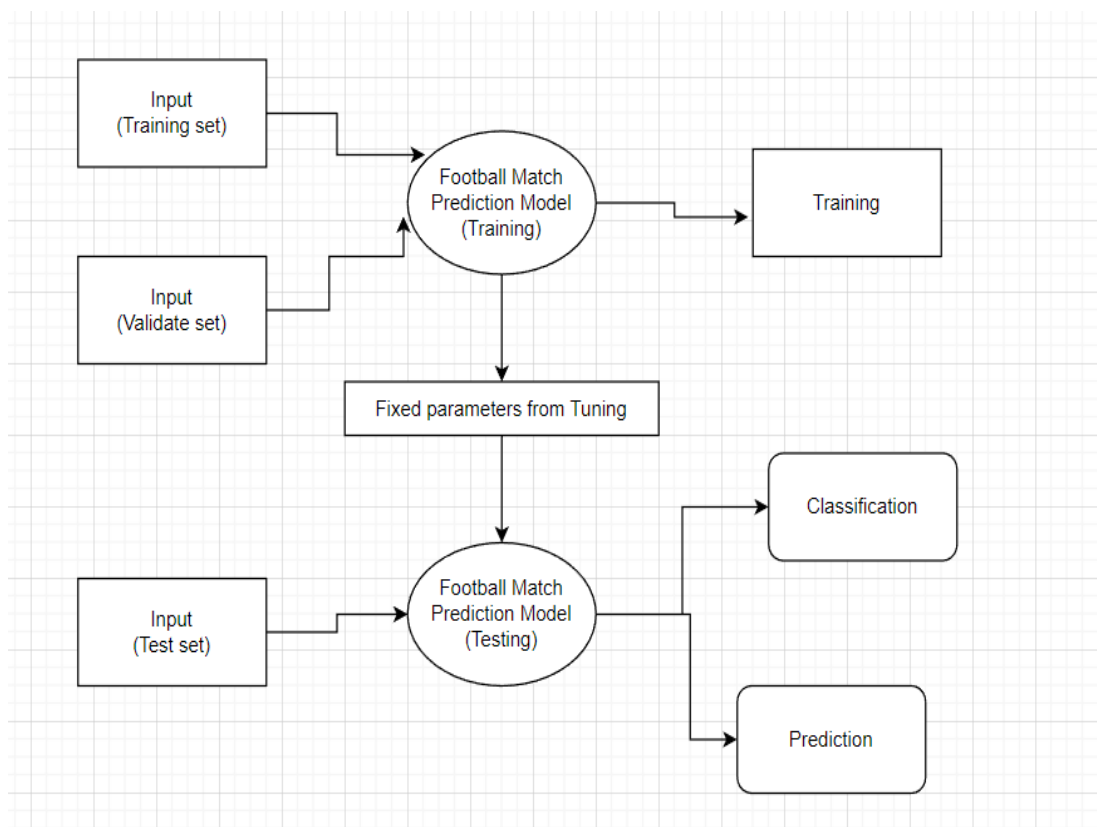


4.2 DATA FLOW DIAGRAM

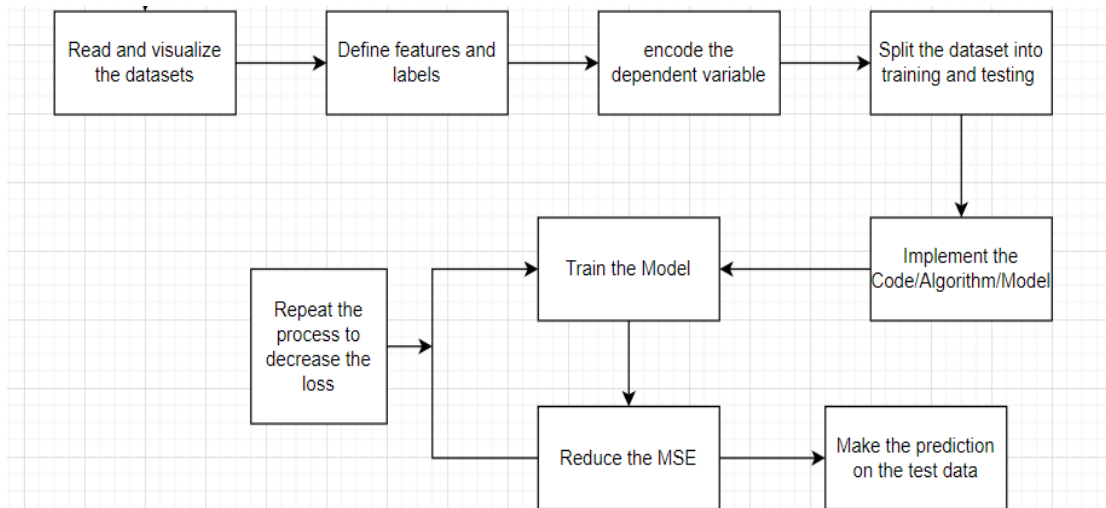
4.2.1 LEVEL - 0



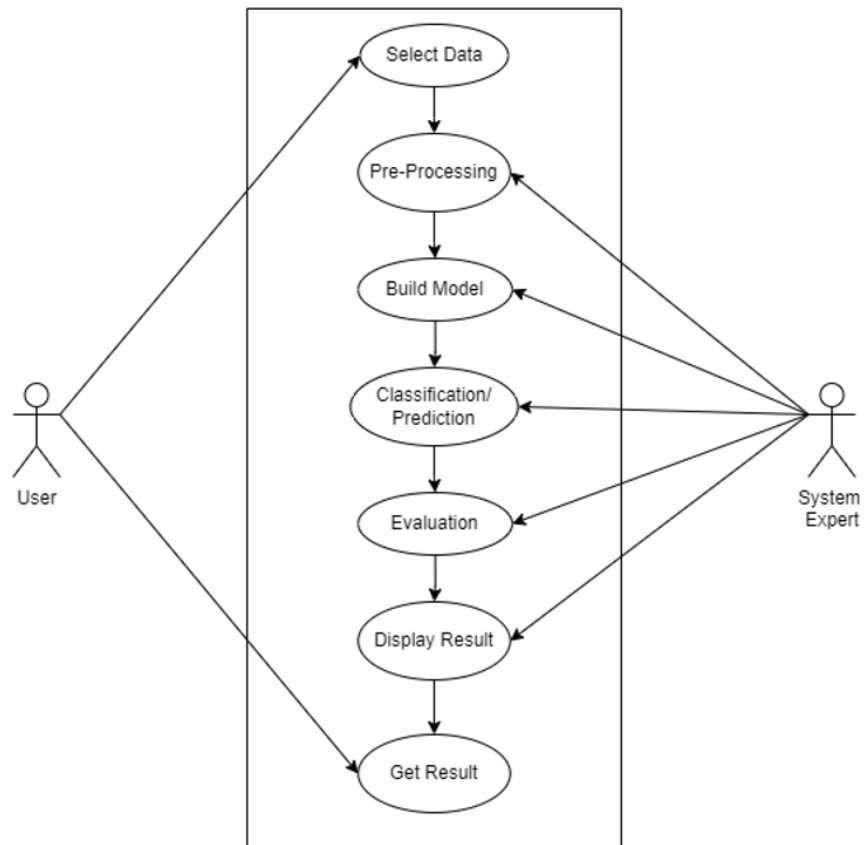
4.2.2 LEVEL - 1



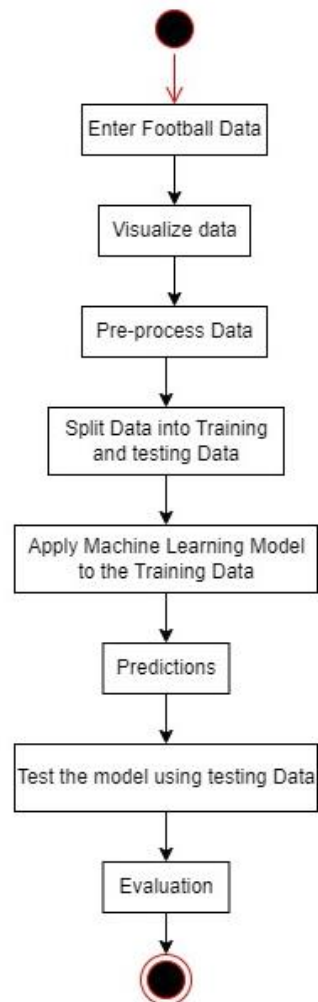
4.2.3 LEVEL - 2



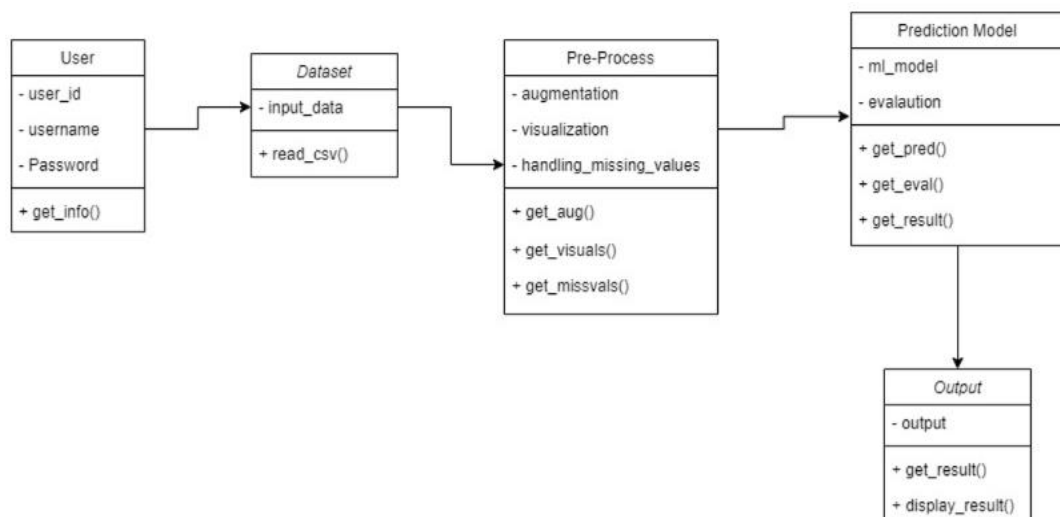
4.3 USE CASE DIAGRAM



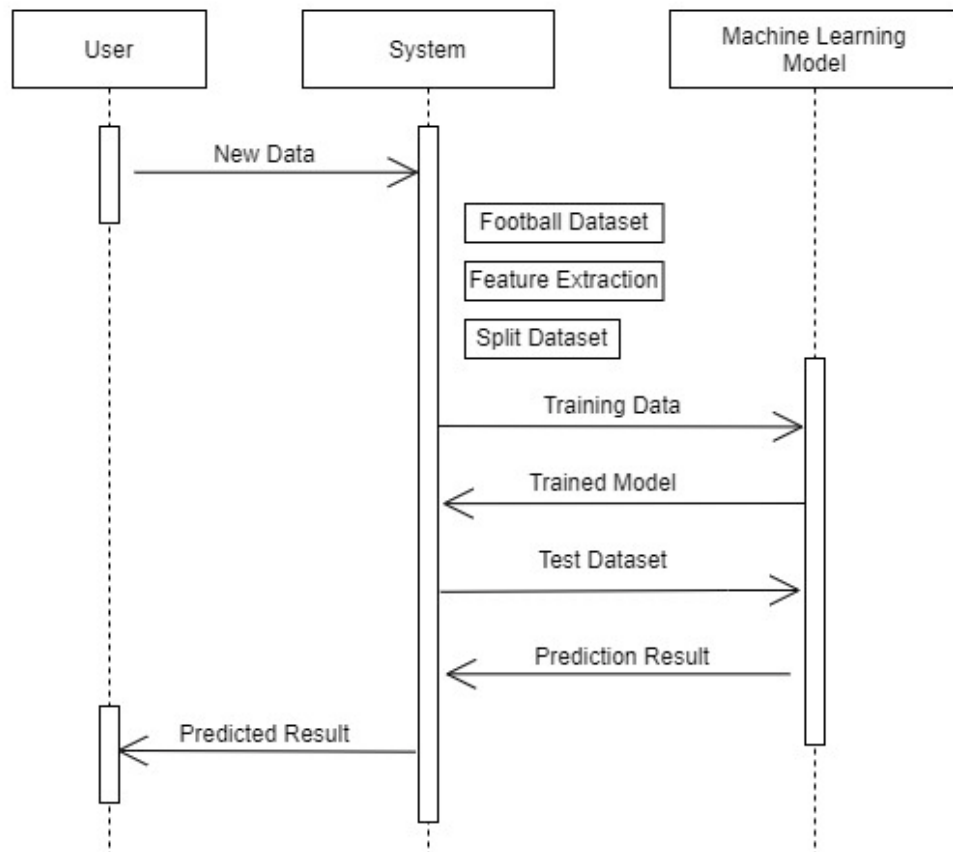
4.4 ACTIVITY DIAGRAM



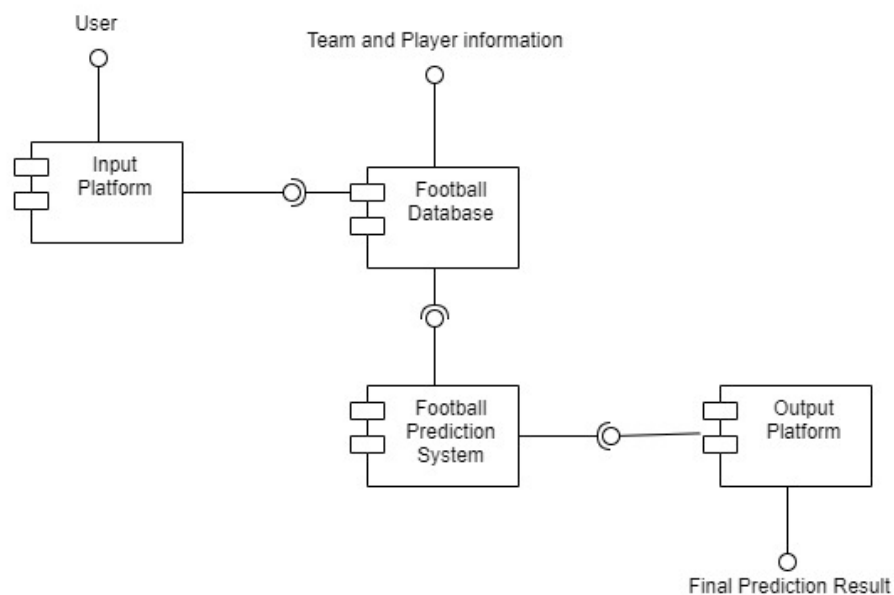
4.5 CLASS DIAGRAM



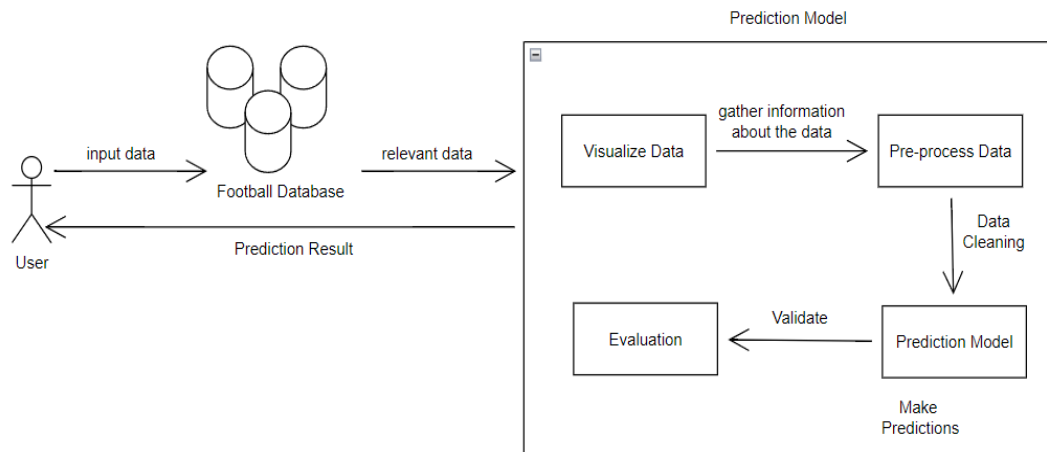
4.6 SEQUENCE DIAGRAM



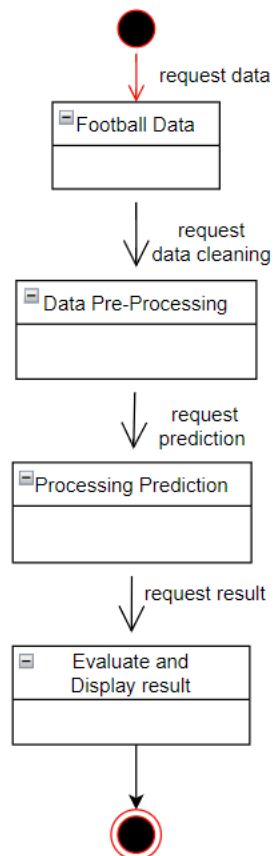
4.7 COMPONENT DIAGRAM



4.8 COLLABORATION DIAGRAM



4.9 STATE CHART DIAGRAM



RESULT ANALYSIS

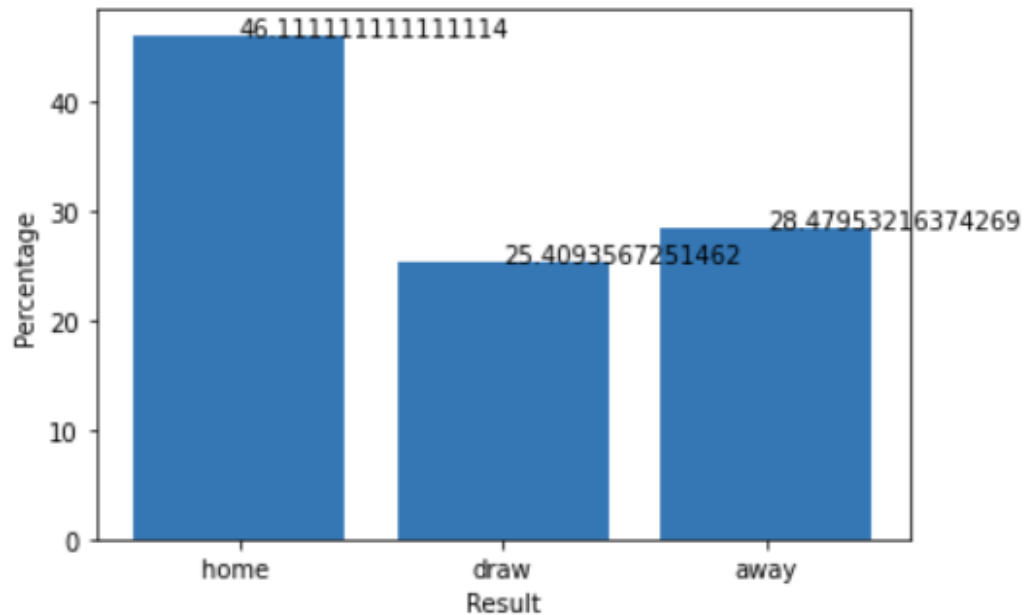


Figure 5.1: The impact of home-away locations on team.

As you can see, home teams have won almost double the times away teams have. This shows that home stadium has an advantage over the result. The fans of the home team are always more in number as compared to away fans. Hence, the support, the noise, and the push from home fans is more.

The fans almost act as a 12th man playing in the field. There is always the fact that the stadium noise affects how players play on that particular day. No matter how all they have been over the period, or what record they have before the match, but the match played on away stadium is always considered to be a tricky match.

Hence, according to the graph, home team win chance is more than Draw or Away Team win chance.

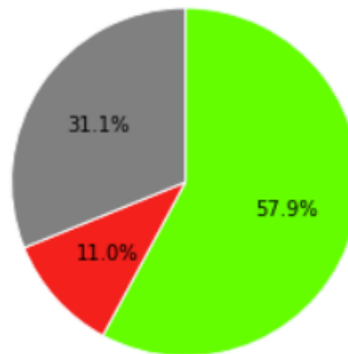
5.1 Home Win Analysis:

725 2 Arsenal Man United 35 1.0 0.0

The above record from the dataset used by us is a match played between Arsenal and Manchester United in the 2nd season, 35th Matchweek. This ended in Arsenal winning the match.

Our system predicts this accurately as shown in figure 5.2:

```
predict_result(2,"Arsenal", "Man United", 35)
```



Home Team Chances: 57.88 %
Draw Chances: 31.14 %
Away Team Chances: 10.99 %

t[135]: 'Arsenal wins'

Figure 5.2

As you can see that our system predicts that Arsenal will win this game with a win percent of 57.88%. Arsenal indeed wins this game.

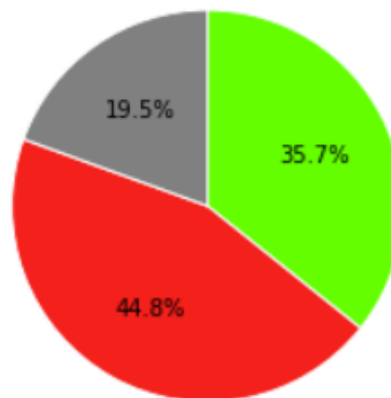
5.2 Away Win Analysis:

2780	8	Burnley	Man City	13	0.0	1.0
------	---	---------	----------	----	-----	-----

The above record from the dataset used by us is a match played between Burnley and Manchester City in the 8th season, 13th Matchweek. This ended in Manchester City winning the match.

Our system predicts this accurately as shown in figure 5.3:

```
predict_result(8,"Burnley", "Man City", 13)
```



Home Team Chances: 35.71 %

Draw Chances: 19.46 %

Away Team Chances: 44.82 %

```
139]: 'Man City wins'
```

Figure 5.3

As you can see that our system predicts that Manchester City will win this game with a win percent of 44.82%. Manchester City indeed wins this game.

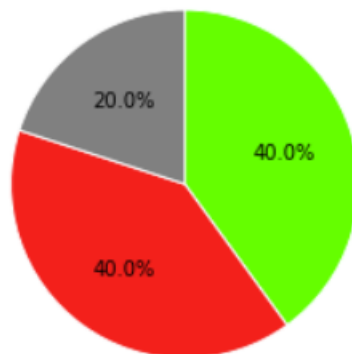
5.3 Draw Analysis:

500	2	Aston Villa	Man United	13	0.5	0.5
-----	---	-------------	------------	----	-----	-----

The above record from the dataset used by us is a match played between Aston Villa and Manchester United in the 2nd season, 13th Matchweek. This match ended in a draw.

Our system predicts this accurately as shown in figure 5.4:

```
predict_result(2,"Aston Villa", "Man United", 13)
```



Home Team Chances: 40.0 %

Draw Chances: 20.0 %

Away Team Chances: 40.0 %

: 'Draw'

Figure 5.4

As you can see that our system predicts that Manchester City and Aston Villa will draw this game with a draw percent of 40.00%. This game indeed ends in a draw.

5.4 Model Accuracy:

In total, 4 models were used in this project:

1. Logistic Regression
2. Naive Bayes
3. Random Forest
4. XGBoost
5. XGBoost - Optimized

The Results were as shown in the Figure 5.5:

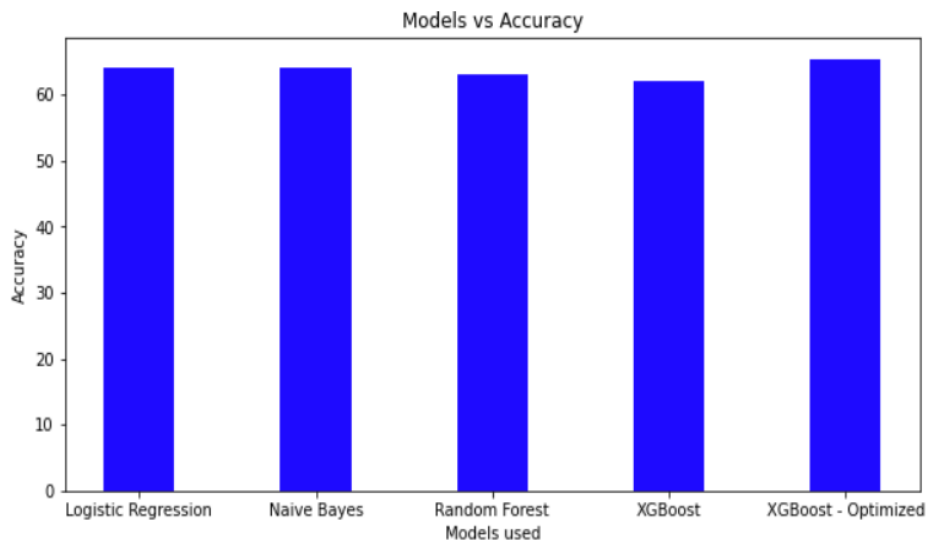


Figure 5.5

XGBoost has the highest accuracy of 65.35% amongst all the models used.

CHAPTER 6

TESTING

6.1 UNIT TESTING

In our project, unit testing could involve writing tests to verify the functionality of the `predict_result()` function. This could include tests to ensure that the function correctly calculates the win/loss/draw percentages, tests to verify that the function generates the expected output for different input values, and tests to check for error conditions (e.g. when there are no matches found for the selected season/home team/away team combination).

6.1.1 IN CASE OF RELEGATION OF A TEAM:

```
predict_result(9,"Hull", "Stoke", 38)
```

The team did not play in the following season

This error is shown in the system when the team is not playing the particular season, as the team is relegated that season. In this case, in the 9th season, this game of Hull City Vs Stoke City does not exist as Hull City was relegated in the previous season. In Figure 6.1 we have shown every match played in the 38th matchweek in the 9th season.

3410	9	Burnley	Bournemouth	38	0.0	1.0
3411	9	Crystal Palace	West Brom	38	1.0	0.0
3412	9	Huddersfield	Arsenal	38	0.0	1.0
3413	9	Liverpool	Brighton	38	1.0	0.0
3414	9	Man United	Watford	38	1.0	0.0
3415	9	Newcastle	Chelsea	38	1.0	0.0
3416	9	Southampton	Man City	38	0.0	1.0
3417	9	Swansea	Stoke	38	0.0	1.0
3418	9	Tottenham	Leicester	38	1.0	0.0
3419	9	West Ham	Everton	38	1.0	0.0

Figure 6.1

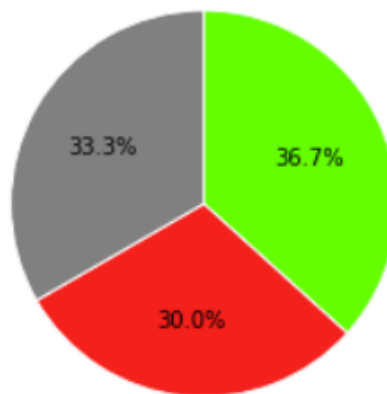
6.1.2 IF THE SEASON INPUT IS WRONG:

```
predict_result(9,"Hull", "Stoke", 12)
```

The team did not play in the following season

This error is shown in the system when the season number is inputted wrong. In this case, in the 9th season, 12th matchweek, the game of Hull City Vs Stoke City does not exist as the actual match took place in the 1st season, 12th matchweek, where Hull City won the match as shown in Figure 6.2:

```
predict_result(1,"Hull", "Stoke", 12)
```



Home Team Chances: 36.67 %
Draw Chances: 33.33 %
Away Team Chances: 30.0 %

```
[163]: 'Hull wins'
```

Figure 6.2

6.1.3 IF THE NAME OF THE TEAM IS ENTERED WRONG:

```
predict_result(1,"H3ull", "St5oke", 12)
```

```
The team did not play in the following season  
The team did not play in the following season
```

Here, you can see both “Hull” and “Stoke” are spelled wrong. Therefore, the error message is shown twice, each for every error made.

CONCLUSION & FUTURE WORK

7.1 CONCLUSION

We are trying to build a model that would provide the highest accuracy with respect to the datasets chosen by us. This model will come in handy to the team analysts and strategists to make their decisions regarding a particular match, by predicting and simulating various scenarios and picking up the best outcome provided amongst them.

7.2 FUTURE WORK

In Future, we would suggest to try different models on the same dataset . This can also be tried on Live data using live inputs and the updates of the match. It will help the tech team and the analysts panel of the respective teams, as this will be more effective to the users.

Finally, many sentimental parameters like fatigue, match sharpness, off-pitch behaviour, general public mentality, injury proneness, relation to the manager of the team, team spirit, confidence and passion towards the team and list goes on. These emotional points can also affect a player in huge way as “stats” isn’t the only way to test a greatness of a player and by implementing all of these parameters can we find more and more optimum solutions.

REFERENCES

- [1] Yoel F. Alfredo, Sani M. Isa, "Football Match Prediction with Tree Based Model Classification", International Journal of Intelligent Systems and Applications(IJISA), Vol.11, No.7, pp.20-28, 2019. DOI:10.5815/ijisa.2019.07.03

- [2] S. N. Maozad, S. Noor Asyikin Mohd Razali, A. Mustapha, A. Nanthamornphong, M. H. Abdul Wahab and N. Razali, "Comparative Analysis for Predicting Football Match Outcomes based on Poisson Models," 2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Prachuap Khiri Khan, Thailand, 2022, pp. 1-4, doi: 10.1109/ECTI-CON54298.2022.9795385.

- [3] Li, H. Analysis on the construction of sports match prediction model using neural network. Soft Comput 24, 8343–8353 (2020). <https://doi.org/10.1007/s00500-020-04823-w>

- [4] Jan Kozak, Szymon Głowania, Heterogeneous ensembles of classifiers in predicting Bundesliga football results, Procedia Computer Science, Volume 192, 2021, Pages 1573-1582, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2021.08.161>.

- [5] Rosli, Che & Saringat, Mohd & Razali, Nazim & Mustapha, Aida. (2018). A Comparative Study of Data Mining Techniques on Football Match Prediction. Journal of Physics: Conference Series. 1020. 012003. 10.1088/1742-6596/1020/1/012003.

- [6] Yash Ajgaonkar, Anagha Patil, Kunal Bhoyar, Jenil Shah, 2021, Prediction of Winning Team using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NTASU – 2020 (Volume 09 – Issue 03)

- [7] M. Vashist, V. Bahl, N. Sengar and A. Goel, "Machine Learning for Football Matches and Tournaments," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 2022, pp. 489-496, doi: 10.1109/COM-IT-CON54601.2022.9850673

- [8] Chen, H. (2019) Neural Network Algorithm in Predicting Football Match Outcome Based on Player Ability Index. *Advances in Physical Education*, 9, 215-222. doi: 10.4236/ape.2019.94015.
- [9] Riccardo Ievoli, Lucio Palazzo, Giancarlo Ragozini, On the use of passing network indicators to predict football outcomes, *Knowledge-Based Systems*, Volume 222, 2021, 106997, ISSN 0950-7051, <https://doi.org/10.1016/j.knosys.2021.106997>
- [10] H R Azhari et al 2018 J. Phys.: Conf. Ser. 1108 012066 DOI 10.1088/1742-6596/1108/1/012066
- [11] C. Pipatchatchawal and S. Phimoltare, "Predicting Football Match Result Using Fusion-based Classification Models," 2021 18th International Joint Conference on Computer Science and Software Engineering (JCSSE), Lampang, Thailand, 2021, pp. 1-6, doi: 10.1109/JCSSE53117.2021.9493837.
- [12] Guan, S., Wang, X. Optimization analysis of football match prediction model based on neural network. *Neural Comput & Applic* 34, 2525–2541 (2022). <https://doi.org/10.1007/s00521-021-05930-x>
- [13] D. Prasetyo and D. Harlili, "Predicting football match results with logistic regression," 2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA), Penang, Malaysia, 2016, pp. 1-5, doi: 10.1109/ICAICTA.2016.7803111.
- [14] Y. W. Syaifudin and P. Puspitaningayu, "Predicting Winner of Football Match Using Analytical Hierarchy Process: An Analysis Based on Previous Matches Data," 2021 International Conference on Data Analytics for Business and Industry (ICDABI), Sakheer, Bahrain, 2021, pp. 47-52, doi: 10.1109/ICDABI53623.2021.9655836.

- [15] J. Almulla and T. Alam, "Machine Learning Models Reveal Key Performance Metrics of Football Players to Win Matches in Qatar Stars League," *IEEE Access*, vol. 8, pp. 213695–213705, 2020, doi: 10.1109/ACCESS.2020.3038601.
- [16] A. A. Azeman, A. Mustapha, N. Razali, A. Nanthamomphong, and M. H. Abd Wahab, "Prediction of football matches results: Decision forest against neural networks," *ECTI-CON 2021 - 2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology: Smart Electrical System and Technology, Proceedings*, pp. 1032–1035, May 2021, doi: 10.1109/ECTI-CON51831.2021.9454789.
- [17] D. Wang, "Clustering and Evolutionary System Analysis of Data Mining Algorithms in the Field of Football," *2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers, IPEC 2022*, pp. 1341–1344, 2022, doi: 10.1109/IPEC54454.2022.9777487.
- [18] S. N. Maozad, S. Noor Asyikin Mohd Razali, A. Mustapha, A. Nanthamomphong, M. H. Abdul Wahab, and N. Razali, "Comparative Analysis for Predicting Football Match Outcomes based on Poisson Models," pp. 1–4, Jun. 2022, doi: 10.1109/ECTI-CON54298.2022.9795385.
- [19] S. Manish., V. Bhagat, and R. M. Pramila, "Prediction of football players performance using machine learning and deep learning algorithms," *2021 2nd International Conference for Emerging Technology, INCET 2021*, May 2021, doi: 10.1109/INCET51464.2021.9456424
- [20] S. K. Nivetha, M. Geetha, R. C. Suganthe, R. M. Prabakaran, S. Madhuvanan, and A. M. Sameer, "A Deep Learning Framework for Football Match Prediction," *2022 International Conference on Computer Communication and Informatics, ICCCI 2022*, 2022, doi: 10.1109/ICCCI54379.2022.9740760.
- [21] Rahman, M.A. A deep learning framework for football match prediction. *SN Appl. Sci.* **2**, 165 (2020). <https://doi.org/10.1007/s42452-019-1821-5>

- [22] N. Danisik, P. Lacko and M. Farkas, "Football Match Prediction Using Players Attributes," 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA), 2018, pp. 201-206, doi: 10.1109/DISA.2018.8490613.
- [23] Qiyun Zhang, Xuyun Zhang, Hongsheng Hu, Caizhong Li, Yinping Lin, Rui Ma, Sports match prediction model for training and exercise using attention-based LSTM network, Digital Communications and Networks, Volume 8, Issue 4, 2022, Pages 508-515, ISSN 2352-8648, <https://doi.org/10.1016/j.dcan.2021.08.008>.
- [24] N. Z. S. T. P. S. U. S. L. K. T., "Prediction of Football Match Score and Decision-Making Process", *IJRITCC*, vol. 6, no. 2, pp. 162–165, Feb. 2018.v[11]
- [25] Hervert-Escobar, L., Hernandez-Gress, N., Matis, T.I. (2018). Bayesian Based Approach Learning for Outcome Prediction of Soccer Matches. In: , et al. Computational Science – ICCS 2018. ICCS 2018. Lecture Notes in Computer Science(), vol 10862. Springer, Cham. https://doi.org/10.1007/978-3-319-93713-7_22
- [26] Rahman, Muhammad & Mustapha, Aida & Fauzi, Rahmat & Razali, Nazim. (2018). Bayesian Approach to Classification of Football Match Outcome. International Journal of Integrated Engineering. 10. 10.30880/ijie.2018.10.06.022.
- [27] T. Korotyeyeva, R. Tushnytsky and V. Kulyk, "Applying Neural Networks to Football Matches Results Forecasting," 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), 2018, pp. 278-282, doi: 10.1109/STC-CSIT.2018.8526756.
- [28] E. Tiwari, P. Sardar and S. Jain, "Football Match Result Prediction Using Neural Networks and Deep Learning," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2020, pp. 229-231, doi: 10.1109/ICRITO48877.2020.9197811.
- [29] L. Hervert-Escobar, T. I. Matis and N. Hernandez-Gress, "Prediction Learning Model for Soccer Matches Outcomes," 2018 Seventeenth Mexican International

Conference on Artificial Intelligence (MICAI), 2018, pp. 63-69, doi: 10.1109/MICAI46078.2018.00018.

[30] Fátima Rodrigues, Ângelo Pinto, Prediction of football match results with Machine Learning, Procedia Computer Science, Volume 204, 2022, Pages 463-470, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2022.08.057>.