**PROJECT REPORT**

**on**

**Project on a Football Match Predicting Machine learning**

**Model**

**(CSE V Semester Mini project )**

**2023-2024**



**Submitted to: Submitted by:**

Ms. Shagun Dasawant Arun Uniyal

(CC-CSE-C-V-Sem) Roll. No:2118319

CSE-C-V-Sem

Session: 2023-2024

**DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY**

**GRAPHIC ERA HILL UNVERSITY, DEHRADUN**

**CERTIFICATE**

**Certified that Mr. Arun Uniyal (Roll No.- 2118319) has developed**

**mini project on “Football Match Predicting Machine learning**

### Model” for the CS V Semester Mini Project Lab in Graphic Era Hill University, Dehradun. The project carried out by Students is their own work as best of my knowledge.

Date:12-01-2024

Ms. Shagun Dasawant

**Class Co-ordinator**

**CSE-C-V-Sem**

(CSE Department)

GEHU Dehradun

**ACKNOWLEDGMENT**

I wish to thank my parents for their continuing support and encouragement. I also wish to thank them for providing me with the opportunity to reach this far in my studies.

I would like to thank particularly my project Co-ordinator Ms Shagun Dasawant.

At last but not the least I greatly indebted to all other persons who directly or indirectly helped me during this work.

**Mr. Arun Uniyal**

**Roll No.- 2118319**

**CSE-C-V-Sem**

**Session: 2023-2024**

**GEHU, Dehradun**

**TABLE OF CONTENTS**

**CHAPTER NO. TITLE PAGE NO.**

**1 Abstract 5**

**2 Introduction 6**

**3 Research Methodology 7**

**4 Evaluation 10**

**5 Conclusion 12**

**ABSTRACT**

This research paper explores the application of logistic regression in developing a machine learning model for predicting football match outcomes. Leveraging a comprehensive dataset encompassing team statistics such opponent team, venue and day at which game is being played and then predicting that if the team is going to win ,draw or lose the match. The implications of this research extend to sports analytics, aiding teams, analysts, and enthusiasts in making more informed decisions based on data-driven insights.

Introduction:

The world of football, particularly the English Premier League, has witnessed an exponential growth in data availability and technological advancements in recent years. This has created an unprecedented opportunity for data-driven decision-making, especially in predicting the outcomes of football matches. In the context of the 2022-23 Premier League season, this project aims to leverage machine learning techniques to predict the results of football matches. Accurate predictions can be of immense value to football enthusiasts, sports bettors, and even clubs and coaches looking to gain a competitive edge. To achieve this, I will explore various algorithms and select the most suitable one based on their performance and characteristics.

Research Methodology:



The liver disease prediction (LDP) method

The LDP method involved in this research are Sample, Explore, Modify, Data preprocessing, Model, Assess and Results. Along with these steps from the SEMMA lifecycle, two more steps, Data preprocessing and Results, are added to this research process.

Sample:

Data collected for this project is from platform Kaggle. The dataset contains the information of premier league teams for the season 2020-21 and then ongoing season 2021-22. After observing the credibility of the dataset from the official data present in the official premier league website, this dataset is use for the project.

Explore:

The second stage is ‘Explore’. Exploring the data stage involves data understanding. This

exploration stage also comprises finding the surprising trends and patterns present in the data to generate . In this stage, dataset was observed for any missing values , correlation between different attributes . This analysis is performed using python.

Modify:

The third stage is ‘Modify’. Modify refers to data transformation . In this study, attributes in the dataset are not in the same format are converted into same integer type as ml model can easily work on numerical data , use that for analysis and henceforth for predictive purpose.

Data Preprocessing:

The fourth stage is Data preprocessing. This data preprocessing refers to cleaning and preparing the data for modelling. It involves replacing the missing data with mean values.

Model:

The 5th stage is the ‘Model’. The modelling stage means applying the selected techniques or the algorithm to the data, here Random forest regression is applied.

Assess:

Assess stage, which is the sixth stage, involves assessing the data by deciding whether the data produced from modelling techniques are reliable and accurate. This stage also evaluates how well the algorithms performed on the data.

Results:

The seventh stage of the proposed LDP method is ‘Results’. The results stage involves presenting the results after assessing the data

EVALUATION:

Evaluation is done by taking all the data of matches before 2022-01-01 as training set while all the matches after 2022-01-01 as test set in order to evaluate how model is working.

After training model on the training dataset and by then evaluating the model in the test set I got:

prediction 0 1

actual

0 157 15

1 48 56

I achieved the precision score of 78.8773%.

actual predicted date team opponent

55 0 0 2022-01-23 Arsenal Burnley

56 1 0 2022-02-10 Arsenal Wolves

57 1 0 2022-02-19 Arsenal Brentford

58 1 1 2022-02-24 Arsenal Wolves

59 1 1 2022-03-06 Arsenal Watford

... ... ... ... ... ...

1312 1 0 2022-03-13 Wolverhampton Wanderers Everton

1313 0 0 2022-03-18 Wolverhampton Wanderers Leeds United

1314 1 0 2022-04-02 Wolverhampton Wanderers Aston Villa

1315 0 0 2022-04-08 Wolverhampton Wanderers Newcastle Utd

1316 0 0 2022-04-24 Wolverhampton Wanderers Burnley

[276 rows x 6 columns]

Conclusion:

In conclusion, this research has demonstrated the efficacy of my machine learning model in predicting football match outcomes, leveraging advanced techniques, particularly Random Forest Regression. Through a meticulous process of data preprocessing, feature engineering, and model training, my approach successfully harnessed key attributes such as opponent team, venue, time, and day of the match to yield accurate and insightful predictions.

My results indicate a significant improvement over baseline models, showcasing the superiority of Random Forest Regression in handling the complexity of football match dynamics.

In conclusion, this study not only contributes to the field of sports analytics but also lays the foundation for the development of more sophisticated and adaptive models capable of unraveling the intricacies of football match outcomes. As I celebrate the success of my current model, the journey towards precision and reliability in football match prediction remains an ongoing and exciting pursuit.