International Journal of Research in Engineering, Science and Management
Volume-3, Issue-2, February-2020
www.ijresm.com | ISSN (Online): 2581-5792

Football Match Prediction Website

Kalyani Ramesh Jadhav[1], Vivita Dias[2], Yadnesh Patil[3], Harshali Rambade[4]
[1,2,3]*Student, Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, India*
[4]*Professor, Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, India*

***Abstract*: Analysis of sports scores and betting is the current trend in market now. Football is one of the most viewed games and people are eager to know which team will win the match even before the match takes place. Hence, football match winning prediction is one of the most researched topics in the market. The part of predicting scores and winning margin of teams is very difficult as football is not an easily predictable game. Legal betting is done not only based on winning, but also based on the scores. Hence, accurate prediction of scores is also a challenging task. We have used Poisson distribution for the prediction of scores and winning team. It is a statistical modelling approach. In this paper, we are predicting the match outcomes of Champions League, by analyzing the past football matches between the teams.**

***Keywords*: Data mining, prediction system.**

1. Introduction

Prediction systems are widely used nowadays in different areas like stock markets, business analysis in shops and companies etc. as well as in sports. Sports' betting is also growing rapidly and hence this makes it one of the most researched topics.

Predicting scores in football matches is difficult as it is not an easily predictable game. We can use machine learning algorithms to predict the chances of winning of teams. So, the main aim of our project is to build a system which can accurately predict the winning, losing and chances of draw between the teams as well as predict the chances of scoring each goal. In this project we have used Poisson distribution for predicting the scores and match outcomes of football match. The model is based on the number of goals scored and conceded by both the teams. Teams which have been high scorers in the past are most likely to score more goals in future. We just need to calculate the average number of goals scored by each team and feed this data to the Poisson model. Our tests have shown that Poisson distribution can be successfully used for calculating the match outcomes.

The project is confined to Champions League matches where the data of the previous matches between both the teams are taken and based on that chances of winning of each team is calculated as well as the probability of scoring each goal is calculated.

2. Literature survey

A. Comparative study of data mining techniques for football prediction

In this research they have compared some of the data mining techniques like decision trees, naïve Bayes, k-means algorithm. This paper used these techniques to predict the match outcomes that is win, lose and draw classes. The data was obtained from English Premiere League matches for three seasons. Their dataset consisted of full time away and home goals, full time result (win, draw, lose), the number of home and away team shots, the number of home and away shots on target and corner. Their experiments proved that decision trees give better accuracy as compared to other algorithms mentioned in the paper. Its accuracy was almost 99.5 %. Bayesian Networks show worst average accuracy around 76 %. [1]

B. A machine learning framework for sports result prediction

They have used artificial neural networks to predict different sports. They have identified the learning methodologies used and the challenges of predicting the sports results. They have given detailed explanation about data understanding, data preparation, feature extraction, modelling, training and testing of model and its evaluation. They have proposed a "SRP-CRISP-DM" framework for sports result prediction [2].

C. Football match winner prediction

In this paper, they have predicted outcomes of English Premiere League based on the historical matches' data gathered from the last 10 seasons and used features like goals scored and conceded, shot ratio etc. They have used logistic regression and used voting algorithm between Naïve Bayes and Random Forest Classifier. Predicting draw cases reduces the accuracy of model. If draw cases are removed its accuracy can be improved up to 65%. Logistic regression fails to predict the draw instances. Hence, the voting algorithm is used to achieve generality [3].

D. Prediction of the FIFA World Cup 2018

In this paper, they have used random forest approach with estimated team ability parameters using ranking methods. They have compared three modelling approaches i.e. Poisson regression method, random forest method and ranking methods [5].

E. Predicting football scores via poison regression model

They have used the Poisson regression model to two national competitions -the 2012-2013 English Premiere League and 2015 Brazilian Football League. It is assumed that the number of goals scored follows Poisson distribution whose average is used to find the attack strength, defense strength and home advantage. It produces satisfactory results. It clearly indicated the winners of EPL and BFL [7].

3. Proposed System

Our system predicts the outcomes of a football match based on the information about the teams which is provided as an input to the trained Poisson Model. The models that we studied only gave winning and losing prediction but the model developed by us predicts winning, losing, draw as well as probability of scoring goals by both the teams. The data gathered for training the model is scraped from different websites like uefa.com, transfermarket.uk, bbc.com. The scraped data is processed to obtain the important attributes from the whole set of data attributes. This input data is given to the prediction model which then computes the probability and the output is shown on the website. Hence, the user does not have to search different websites to get the information. Our system offers all in one information about the teams as well. We have developed a website, so that the user can easily interact with the system. The website is developed using ASP.net, HTML and CSS. The website has an admin page where the admin can update the details of upcoming matches as well as scrape more data of matches and store it in the database. We have used Structured Query Language (SQL) for implementing the database. The user can use the website to get insight about the winning chances of both the teams and chances of draw between them. User can also look for other important information like top scorers of matches and other player statistics. They will also get information about most assists, clean sheets, etc.

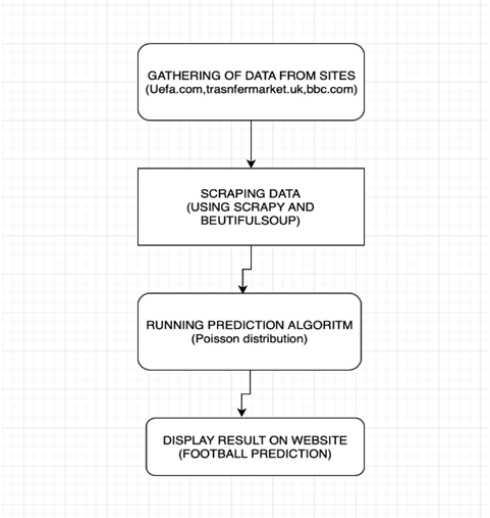

Fig. 1. Proposed system

A. UML Diagrams

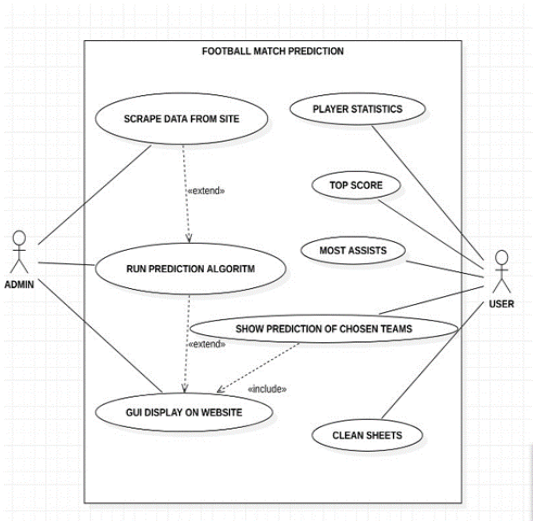

Fig. 2. Use case diagram

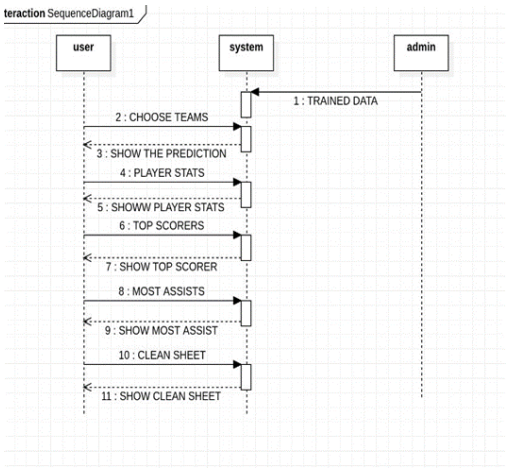

Fig. 3. Sequence diagram

B. Data scraping of football matches

Data scraping, also known as web scraping, is the process of obtaining information from a website into a spreadsheet or local file saved on your system. Here, we are using data scraping to extract information (such as player statistics, home score and away scores of teams, shots, most assists and clean sheets). We are using Beautiful Soup and Scrapy. Beautiful Soup is a python package used for parsing HTML and XML documents [6]. Scrapy is another open source web- crawling framework. The scraped data is, then obtained in JSON format which is then cleaned to obtain only useful information needed for prediction.

C. Display of results

When the user selects the names of two teams, then the data of those teams is extracted from the database and is fed to our prediction model which then computes the results and displays it to the user on the website

D. Prediction algorithm

The model is based on the number of goals scored and

conceded by the teams. It is a discrete probability distribution that describes the probability of events within a specific period of time with a known average rate of occurrence.

Formula used for calculating the average number of scores:

- Average goals scored home = Season total goals scored at home / number of games (in season)
- Average goals scored away = Season total goals scored away / number of games (in season)

We also calculate average goals conceded at home and away by just inverting the above numbers.

4. Conclusion

Our proposed system enables the user to gain insights about the winning and losing chances of both the teams before the match which can be used for betting as well. It gives detailed prediction of the score-line as the betting also takes place based on the goals scored by each team. According to our tests, our model gives quite accurate results and our system provides a user-friendly interface which can be used by anyone. Our system can also be used by football enthusiasts to analyze the strong and weak teams.

References

[1] Che Mohammad Firdaus Che Mohamad Rosli, Mohd Zainuri Saringat, Nazim Razali, Aida Mustapha, "Comparative Study of Data Mining Techniques on Football Match Prediction"
[2] Rory P. Bunker, Fadi Thabtah, "A Machine Learning Framework for Sports Prediction", September 2017.
[3] Saurabh Vaidya, Harshal Sanghavi, Kushal Gevaria, "Football Match Winner Prediction", November 2016.
[4] Aditya Srinivas Timmaraju, Aditya Palnitkar, Vikesh Khanna, "Game On! Predicting English Premiere League Match Outcomes"
[5] Andreas Groll, Christophe Ley, Gunther Schauberger, Hans Van Eetvelde, "Prediction of the FIFA World Cup 2018- A Random Forest Approach with an emphasis on estimated team ability parameters.", June 2018.
[6] https://en.wikipedia.org/wiki/Beautiful_Soup_(HTML_parser)
[7] Erlandson F. Saraiva, Adriano Suzuki, Ciro A. O. Filho, Francisco Louzada, "Predicting Football Prediction Via Poisson Regression Model: Applications to National Football League", July 2016.