Prediction of Football Match Score and Decision Making Process

Nilay Zaveri
IT-SAKEC
Mumbai, India.
nilay.zaveri@sakec.ac.in

Utkarsh Shah
IT-SAKEC
Mumbai, India.
utkarsh.shah@sakec.ac.in

Shubham Tiwari
IT-SAKEC
Mumbai, India.
shubham.tiwari@sakec.ac.in

Pramila Shinde
Assistant Professor
IT-SAKEC
Mumbai, India.
pramila.shinde@sakec.ac.in

Lalit Kumar Teli
IT-SAKEC
Mumbai, India.
lalit.teli@sakec.ac.in

Abstract— Football isone of the most popular sport internationally. With that amount of popularity and fame, it has gained a lot of attention from the technical field. Prediction of match results has been in demand for various purposes including team management, analysis and betting. We have come up with a solution using machine learning algorithms that can fulfil all the current needs in football match prediction. The implementation only includes teams from Spanish La Liga over the last 5 seasons. We have predicted the outcome of matches between Home Team and Away Team which would include the final score, the starting 11 players, the substitutes and the names of probable goal scorers. For the purpose of analysis, we have provided the stats of players and teams referring to the FIFA 18 game database as well as their actual career stats. We have also provided the analysis of strength, weakness and tactics of players and teams. Finally, for decision making purposes, we would make a system that can analyse the Home and Away team and then suggest the tactics to the user for their team that can maximize their winning chances.

_____*****_____

I. INTRODUCTION

Today, the team managers,fans and analysts give a prediction about who is going to win the match. Nowadays, different statistics are used for the prediction of football match result. The prediction of a football match is done by using some previous data. The prediction of the match depends on variables like Team stats, Player stats, Historical data etc.which is used by managers and club directors to decide who is going to win the match and what is needed to win the match. Football result prediction has gained lots of popularity in recent years. Previously, Knowledge discovery in databases (KDD) was used to develop a football match result predictive model by gathering the features that affect the outcome of football matches. Data mining techniques have also been used in thepast.In existing systems, they have analysed two approaches used in football match result prediction, statistical and machine learning. For team management, the manager and the other coaching staff members need to know about their own team as well as the opponent team in detail. They should be able to identify strengths and weaknesses of different teams and accordingly prepare for the matches. The management should also be able to monitor a match in real time to take a decision during a match. For all these purposes, the prediction application should be able to suggest ideas based upon the provided conditions. We have used multiple machine learning algorithms wherein Logistic Regression gave us the best accuracy. Logistic Regression is used when the dependent variable is categorical. In our case, the dependent variable has 3 possible outcomes (Home Win, Away Win and Draw) and hence, multinomial regression has been used in a one vs all mode. This method'smain advantage is that it's very simple to explain the relationship between output variable and input. It allows one to say that the presence of a risk factor increases the probability of a given outcome by a specific percentage. We proposed a model of football match prediction by using the data of every match in La Liga for the last 5 years and training the algorithm based on that data. The algorithm learns from training data to form different rules and pattern and based on that they make a decision on some new data. With every input, the algorithm keeps on learning by using the feedback.

The software predicts the match result, score, starting 11, substitutes and names of goal scorers. These results can be used by the management staff while team selection or by the fans playing fantasy leagues in order to build an efficient team or for the purpose of betting. This software is different from the other software as it uses the machine learning algorithms along with a little statistical approach to improve the accuracy. Another difference is that it can be used by all types of users (fans, management staff and bet placers) because it provides a wide range of functionalities rather than just prediction of Win, Lose or Draw.

A. AIM OF THE PROJECT

The software product "PREDICTION OF FOOTBALL MATCH SCORE &DECISION MAKING PROCESS" will be

a system that provides prediction of match result, squad, substitutes, goal scorers, team analysis, player analysis, player evaluation and decision support to improve the winning chances against any given team.

B. SCOPE OF THE PROJECT

This software will improve the efficiency and will be able to address all the needs of the different users. It will help team managementto know about their own team as well as the opponent team in detail. They'll be able to identify strengths and weaknesses of different teams and accordingly prepare for the matches. The management also is able to monitor a match in real time to take a decision during a match.

II.   LITERATURE SURVEY

An Improved Prediction System for Football a Match Result

Vol. 04, Issue 12 (December 2014)

They have used an approach of KDD using Logistic Regression and ANN. They have performed operations using RapidMiner tool. The application only contains a data of 110 matches of the 2014-2015 EPL and have predicted the results of the further matches of the same season.The data used is very specific to a particular season and the performance of different teams, players and coaches in the starting matches of the season. They have been able to reach an accuracy of 93%. The drawback of the system is that it can only predict Win, Lose or Draw and only for a particular season. Since, there is always a major change when a new season begins, the accuracy of the model for a fresh season can be highly affected. [1]

Football Result Prediction with Bayesian Network in Spanish League-Barcelona Team

Vol. 5, No. 5, October 2013

They have used Bayesian Networks for classification. The system is set for only 1 team (FC Barcelona) and has a vast number of real-time parameters (Psychological and non-psychological factors). Every match is considered separately since all the factors can change with time.The accuracy of the model was 92%. The drawback of the system is that it can only predict for a particular team and needs data input for each match before prediction. Also, it can only predict win, lose or draw. [2]

Predicting Football Match Results with Logistic Regression

They have used Logistic Regression algorithm for prediction of results of 1 season based on the previous 5 seasons. They are only predicting Home Win or Away Win. They have obtained an accuracy of 69.5%. The drawback of the system is that it only predicts the probability for Home Win. If the probability is more than 50% then it is specified as a Home Win otherwise

as an Away Win. The factor that matches can result to a Draw hasn't been considered. [3]

Predicting football scores using machine learning techniques

MIPRO 2011, MAY 23, 2011, Opatija, Croatia

They have implemented the following algorithms:

- Naive Bayes
- Bayesian networks
- LogitBoost
- The k-nearest neighbors algorithm
- Random forest
- Artificial neural networks.

Using those algorithms, they have predicted Home Win, Away Win and Draw. The best accuracy that they could reach was 65%. 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. [4]

A novel way to Soccer Match Prediction

They have used multiple machine learning algorithms for the prediction of Home Win, Away Win and Draw. The system uses the data from FIFA 15 game and also Match History database of EPL season 14-15. The training and testing is done on the same season database and the best accuracy obtained is 80%. The drawback of the system is the fact that it only considers 1 season for training and the testing is done on the same.[5]

III.   PROBLEM DEFINITION

Prediction of the result(Home Win, Away Win or Draw) of a football match in La Liga by specifying the Home and Away team names. Also predicting the exact score-line of the match along with the names of the probable goal-scorers and also other football match attributes like shots, shots-on-target and corners. Finally, include a decision suggestion system for the football team managers to improve the team's performance. The decision support system would suggest team selection, tactic selection, analysis of opponent team, player evaluation etc.

IV.   DESIGN & IMPLEMENTATION

A. Database Design

Five databases have been used, each for a different application:

- The Match History Database for La Liga 12-13 to La Liga 16-17 (5 years) has been used which contains 12 different attributes.
- Also, a Team Vs Team database has been used which contains the records for all the teams against all otherteams in the last 5 seasons in La Liga. These 2 tables

___

are together used forprediction of match result (Home Win, Away Win or Draw).

- Goals history database contain several attributes related to goals, shots, shots on target etc. of the last 5 seasons and is used for the prediction of final scores for each team.
- Player Stats database has been taken from "FIFA 18" game which is used for Player Analysis, Player Evaluation and for selection of Starting 11 players along with the substitutes. It includes 46 different attributes for each player. The database also helps for prediction of goal-scorers.
- Team Stats database has been taken from "FIFA 18" game which will be used for Team Analysis and Suggestion system.

B. Implementation

The system has been implemented using several machine learning algorithms (Logistic Regression, Random Forest, Artificial Neural Network, Linear SVM and Naïve Bayes) using Python Sci-Kit Learn library. The training and testing is done on the complete dataset using Cross-Validation function with cv = 10. The main part of our implementation was concentrated on database pre-processing and attribute selection.

The Match History Database includes 12 attributes: Home Team ID, Home Shots, Home Shots on Target, Home Corners, Home Yellow Cards, Home Red Cards and similar 6 attributes for Away team. A total of 29 teams have participated in La Liga between 2012-2017. These teams have been assigned an ID from 1-29 based on their performance in the 5 seasons where the best team ID starts from 1.

As the 13th attribute we used the Home Team's win percentage against the given Away Team, which was obtained from the Team Vs Team Database. It proved to be successful by increasing the accuracy of the model by almost 8%. This idea was taken from the fact that every team has a playing style and also has rivals. So, even though a team is better than the other team in everything, but due to the rivalry or due to their respective playing styles, the result can be a loss. Hence, the inclusion of Team ID and Head to Head win percentage was successful.

Further, we used the Goals History database to predict the exact score of both the teams at the end of the match using Logistic Regression and the accuracy was found out to be 69.87%. After this, we used the statistical and random simulation approach to find the goal scorers. The team and player analysis were displayed directly from the databases and represented graphically as in [5].

Later we would make the player evaluation system to find whether a new player can fit in a squad and also their transfer value. Finally, the decision support system will be

implemented using a statistical approach along KDD to find the strength and weakness of all the teams based on which suggestions will be made to improve the winning chances.
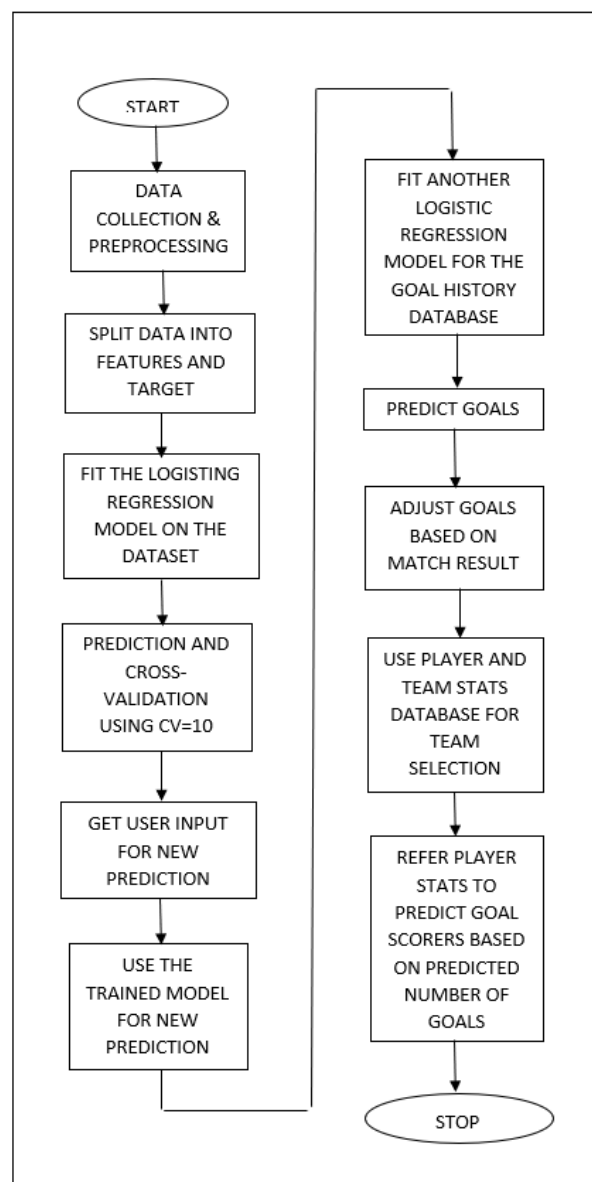

Figure1. Project Flow Diagram

C. Results Obtained

TABLE 1. ALGORITHM ACCURACY TABLE

	LR	RF	ANN	Linear SVM	NB
Match History Database	63.94%	61.53%	63.1%	58.25%	58.63%
Match History Database + Team Vs Team Database	71.63%	69.9%	69.2%	66.95%	63.57%

___

LR = Logistic Regression

RF = Random Forest

ANN = Artificial Neural Network

NB = Naïve Bayes

SVM=Support Vector Machine

The Accuracy specified in the table is the average of the accuracy for Home Win, Away Win and Draw classifications.

D. Regression Coefficients

TABLE 2. REGRESSION COEFFICIENTS TABLE

	Home Win	Away Win	Draw
HT ID	-0.027	-0.008	0.020
AT ID	0.009	-0.015	-0.004
HT Win percentage vs AT	4.613	-4.836	-0.630
HT Shots	-0.112	0.072	0.030
AT Shots	0.053	-0.089	0.024
HT Shots on Target	0.449	-0.275	-0.193
AT Shots on Target	-0.347	0.560	-0.193
HT Corners	-0.155	0.083	0.060
AT Corners	0.089	-0.117	0.016
HT Yellow Cards	-0.036	0.027	0.034
AT Yellow Cards	-0.087	0.017	0.104
HT Red Cards	-0.650	0.435	0.128
AT Red Cards	0.489	-0.722	0.093

HT: Home Team

AT: Away Team

V. CONCLUSION

We implemented the model using different machine learning algorithms and were able to reach the accuracy of 71.63% with Logistic Regression on the Match History Database of 5 seasons along with the Team Vs Team Database.Referring to related works, there are a lot of improvements that can be made to our system for improving the accuracy of both, match result prediction as well as goal prediction. Also, the two predictions can be later combined to further improve the model. We can also use the Machine Learning approach for the Goal Scorer prediction.

VI. REFERENCES

[1] Igiri,Chinwe Peace, Nwachukwu, Enoch Okechukwu,"An Improved Prediction Systemfor Football a Match Result" IOSR Journal of Engineering (IOSRJEN), Vol. 04, PP 12-20.

[2] FarzinOwramipur, ParinazEskandarian,and FaezehSadatMozneb,"Football ResultPrediction with Bayesian Network in Spanish LeagueBarcelona Team" InternationalJournal of Computer Theory and Engineering, Vol. 5, No. 5, October 2013.

[3] Darwin Prasetio, Dr.Harlili,"Predicting Football MatchResults with LogisticRegression", Institute Teknologi Bandung, 2016.

[4] Josip Hucaljuk, AlenRakipović,"Predicting football scores using machine learningtechniques", University of Zagreb, Faculty of Electrical Engineering and Computing Unska 3, May 23-27, 2011.

[5] Jongho Shin, Robert Gasparyan, "A novel way to Soccer Match Prediction".