International Journal of Future Generation Communication and Networking
Vol. 13, No. 2s, (2020), pp. 247–253

Study of Various Machine Learning Algorithms to Predict the Outcomes of Soccer Matches

Shubham Rane[1], Akshay Chandgude[2], Ritesh Kumar[3], Rohit Kadam[4]

[1, 3, 4] *UG Student, Department of Electronics and Telecommunication, SKNCOE, Pune, India*
[2]*Assistant Professor, Department of Electronics and Telecommunication, SKNCOE, Pune, India*

[1]*raneshubham162@gmail.com*
[2]*akshaychandgude9@gmail.com*
[3]*ritesh.03k@gmail.com*
[4]*kadamrohit9595@gmail.com*

Abstract

Machine Learning is one of the smart procedures that have demonstrated promising outcomes in the spaces of arrangement and expectation. One of the growing zones requiring great prescient precision is sport forecast, because of the huge financial sums engaged with wagering. Likewise, club chiefs and proprietors are taking a stab at arrangement models so they can comprehend, and plan systems expected to win matches. These models depend on various elements engaged with the games, for example, the aftereffects of verifiable matches, player execution markers, and resistance data. Machine Learning allows us to gain insight into data using which we aim to cover feature extraction for Predicting outcome of soccer matches using machine learning to gain insight. The system will be performing our analysis based on our featured dataset and implement multiple classification algorithms such as support SVM.

Keywords *- Machine Learning (ML), Football result prediction, Soccer, SVM (Support Vector Machine), Premier League.*

I. INTRODUCTION

Sports are extremely well known and boundless type of investing free energy, being healthy and procure cash. Game group administrations are contributing tremendous measures of cash into improvement of their players and mechanical hardware and offices. Collection of information about players and the matches empowers experts to examine past outcomes of the entire group as a unit to conclude effect of different impacts in the group's results. Domain of prescient investigation is additionally outstanding methodology which is intently associated with sports. For whatever length of time that game is unpredictable movement performed by human beings, there is typically an irrelevant factor of haphazardness and different viewpoints such as weather, mind of players or effect of media and fans. That is the reason the games pre-expressions have consistently been trying for the explores and examiners and football (soccer) is being viewed as such a game. When all is said in done, there are two fields, where pre-styles could be applied. Betting organizations dealers are utilizing measurements to define betting chances for the various groups, player and matches to acquire cash for their organizations. Then again, there are individuals, who are attempting to beat these odds and win cash for fruitful mixes of tips of single matches. Even though this is considered as risky action, it shows like an intriguing scientific problem.

ISSN: 2233-7857 IJFGCN

Each round of football keeps going for 90minutes (+ stoppage time) two rival groups of 11 players each safeguard objectives at far edges of a field having goal lines at each end, with focuses being scored predominantly via conveying the ball over the rival's goal line and by place-kicking or drop-kicking the ball over the crossbar between the operation opponent's objective posts. There are 3 potential results of the game host group victory, away group win or draw. Our work is also based on predicting such results of the football matches in order to build reliable model capable of containing knowledge that could be used for optimization of composition of the football team.

II. LITERATURE SURVAY

[1] Maral Haghighat, Hamid Rastegari and Nasim Nourafza proposed various answers for takeout difficulties. For example, forecast precision can be improved using AI and information mining procedures that have not been utilized in this field however have yielded great outcomes in different fields. Use of half and half calculations can likewise help forecast exactness. Besides, including various highlights, for example, player execution will add to progressively precise forecasts. Then again, a thorough dataset can be gathered by the assistance of a gathering of specialists in every game field. So as to give the opportunity to examinations between various investigations, specialists are prescribed to gather information from substantial associations (for example NBA).

[2] Albina Yezus, Machine learning techniques was been applied to various fields, including sports. On the case of English Premier League, it is demonstrated that it is conceivable to discover a classifier that predicts the result of soccer matches with the accuracy of over 60%. Be that as it may, there was still a ton of work to be done and they said research will be continued.

[3] Brianne Boldrin, proposed system where Football is one of the most troublesome games to foresee yet as of late there has been fast development in the territory of anticipating football results through measurable displaying. The most viewed and most gainful football alliance in the word is the English Premier League. Communicate to more than 643 million homes and to 4.7 billion individuals, the Premier League is England's top class with specific spots. They proposed a model after a Poisson dispersion to anticipate the outcomes and the best wagering choices of the 2016-17 Premier League.

[4] Norbert Danisik, Peter Lacko, Michal Farkas proposed plan that depended on the dataset comprising of both match history and player qualities information. These characteristics they are picked up from the computer game called FIFA and they are joined with the information highlights gathered in the genuine football matches. Our iterative way to deal with the work drove us to create 3 primary model structures in generally speaking. At that point they efficiently attempted different setups with cross-approval to deliver the most target results. The best performing was the LSTM relapse model, which achieved normal forecast precision of 52.479\%. This is a promising outcome, which is similar with the distributed best in class arrangements. Results demonstrated both appropriateness of utilization of the LSTM neural systems for such a prescient errand and the positive effect of utilizing the computer game information so as to determine genuine, non-virtual issues.

[5] Siddhesh Sathe, Darshan Kasat, Neha Kulkarni, proposed that their system has Best performing algorithm which was SVM having accuracy of 0.599 followed by Naïve Bayes of 0.55which is better than accuracy of 0.52 of leading BBC analyst Mark Lawrenson[12] and betting organization Pinnacle Sports in which had accuracy of 0.55 which is equivalent to that obtained by naïve Bayes. Random forest with accuracy of 0.50 had lowest accuracy. This accuracy can be further improved by adding more relevant features developing models which take into consider even broader aspects of football.

ISSN: 2233-7857 IJFGCN

[6] Ben Ulmer and Matthew Fernandez "Predicting Soccer Match Results in the English Premier League",2017, proposed system to predict the results of soccer matches in the English Premier League (EPL) using artificial intelligence and machine learning algorithms. From historical data we created a feature set that includes game day data and current team performance (form). Using feature data, they created five different classifiers: Linear from stochastic gradient descent, Naive Bayes, Hidden Markov Model, Support Vector Machine (SVM), and Random Forest. Their prediction was one of three classes for each game: win, draw, or loss. Their best error rates were with our Linear classifier (.48), Random Forest (.50), and SVM (.50) and their error analysis focused on improving hyper parameters and class imbalance.

[7] Tim van der Zaan, "Predicting the outcome of soccer matches in order to make money with betting", Business Analytics and Quantitative Marketing, Erasmus University Rotterdam,2017 One of the main goals their system was to come up with prediction models that accurately predict the outcome of soccer matches. In addition, it is attempted to construct a betting strategy that is able to defeat the bookmakers and generate profit from betting on
soccer matches.

[8] Sumit Shrestha,"Premier League Game Result Prediction", Tribhuvan University,2016.Their application estimates the prediction of the result of premier league game between two teams as win, lose or draw of the game. Their result was overall satisfactory with 47% of the accuracy of the prediction using Back Propagation algorithm.

[9] Chinthey Peace and Nwachukwu, Enoch Okechukwu ,"An Improved Pre- diction System for Football a Match Result", IOSR Journal of Engineering (IOSRJEN), Vol. 04, Issue 12 (December 2014) The improved football result Prediction System explores the use of machine learning techniques in the framework of Knowledge Discovery in Database. Their research was driven by the overwhelming increase in the pool of avail- able sports data in English premier league. The datasets collected was successfully implemented using data mining technique in different aspects of the work. In many instances, predicting the results of sporting procedures has always been a challenging and rewarding venture, therefore forecasting problem provides a growing need to conduct research in this area. Sports outcomes predictive techniques arise, and this motivates the need to find more valuable datasets to improve the prediction accuracy and make precise decisions at key. Past comprehensive statistical data has been kept assisting English premier league games and other sporting events. Both players and teams' present varying forms of these statistical facts kept as data season in and off-season. As the dataset set grows with the EPL games, it has become the preferred test platform. This pool of information will keep motivating different groups, ranging from public, statisticians and sports enthusiasts to discover embedded knowledge in it.

[10] Albina Yezus, "Predicting outcome of soccer matches using machine learning", Saint Petersburg State University Mathematics and Mechanics Faculty,2014 Machine learning methods was been applied to different fields, including sports. On the example of English Premier League, it is shown that it is possible to find a classifier that predicts the outcome of soccer matches with the precision of more than 60%. However, there was still a lot of work to be done and they said research will be proceeded.

[11] Jongho Shin and Robert Gasparyan, "A novel way to Soccer Match Prediction", semantic scholar ,2014They demonstrate an alternative source for curated data: video games. Video games are often overlooked due to its origin. However, video games have come a long way since Pac Man and Frogger and have created phenomenally accurate simulations of the real world, which

can only be done through very intensive data collection. This data can be used in machine learning projects to make predictions in the real world with very accurate results. Of course, this would be made easier if the video game industry shared this information in public domain. FIFA 2015 has done a great job simulating the action of soccer players, but it is likely that they have used more than 33 features, which are presented in the game, to accomplish this.

[12] Maral Haghighat, Hamid Rastegari and Nasim Nourafza, "A Review of Data Mining Techniques for Result Prediction in Sports", ACSIJ Advances in Computer Science: an International Journal, Vol. 2, Issue 5, No.6 , November 2013 Suggested a number of solutions to eliminate challenges. For instance, prediction accuracy can be improved using machine learning and data mining techniques that have not been used in this field but have yielded good results in other fields. Application of hybrid algorithms can also boost prediction accuracy. Moreover, including different features such as player performance will contribute to more accurate predictions. On the other hand, a comprehensive dataset can be collected by the help of a group of experts in each sports field. In order to provide the chance for comparisons between different studies, researchers are recommended to collect data from valid leagues (e.g. NBA).

III. System Architecture

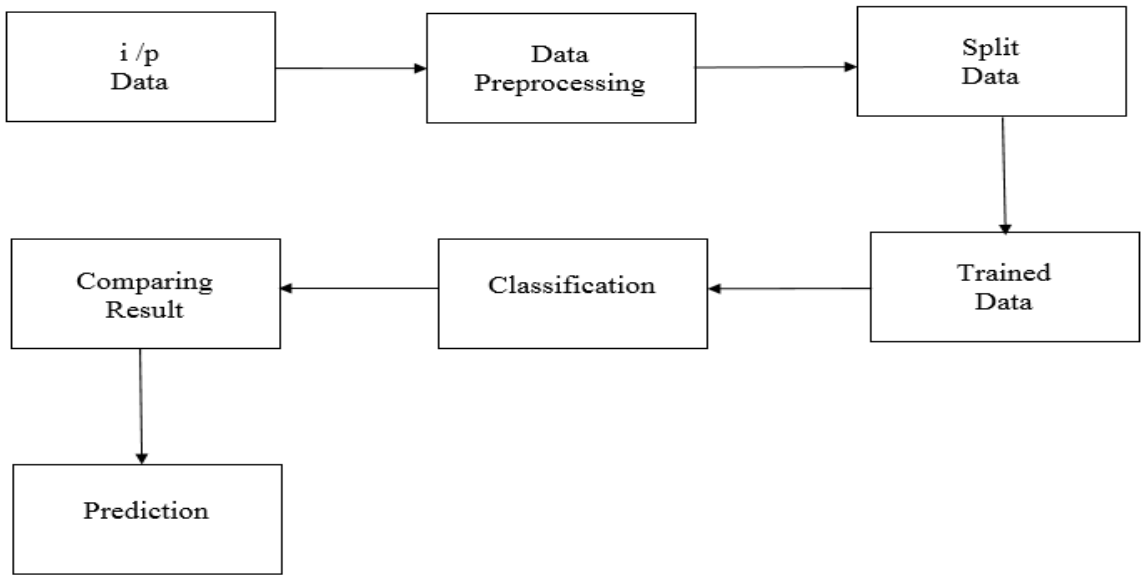

Fig 1: System Architecture

IV. WORKING

We train the final data set on various machine learning classifiers. We compare the performances of each classifier and choose the one that returns the best result. Then, we optimize the classifier that yields the best result to further enhance the model accuracy in making predictions.

A. Dataset Description -

We are going to be predicting match outcomes using data from past games for a few seasons. The data various attributes per season regarding the home team, the away team, the venue, scores, to name a few. We filtered these attributes into a final list which proved to be the most influential for predicting the outcome.

B. Pre-Processing –

In data set, obtained several attributes from each season. A lot of these features are pretty much unnecessary for making outcome predictions. Hence, our primary task is to clean the data to only retain the features or attributes most. We calculate the Scatter Matrix to observe how much one attribute affects another set and their correlations. This will help us pick the most influential features that we want to use to build our new data set.

C. Data Splitting

Once we finish building our new set of crucial attributes, we split the data into training and testing data.

V. SUPPORT VECTOR MACHINE

Support vector machines are models in machine learning that is useful for regression analysis and classification tasks. we map each data item as a point in a space of n-dimensions (n being the number of features) in which each feature-value corresponds to a co-ordinate. the target is to obtain a hyperplane that classifies all training vectors into two classes. the finest choice is the hyper-plane that leaves the maximum margin from both the classes

Case 1: Consider the case with data from 2 different classes. Now, we wish to find the best hyperplane which can separate the two classes. To find which hyperplane best suit this use case. In SVM, we try to maximize the distance between hyperplane nearest data point. This is known as margin.

Case 2: Consider the case with data from 2 different classes. Now, we wish to find the best hyperplane which can separate two classes. As data of each class is distributed either on left or right. Our motive is to select hyperplane which can separate the classes with maximum margin. In this case, all the decision boundaries are separating classes but only 1st decision boundary is showing maximum margin between triangle circles.

Case 3: Consider the case with data from 2 different classes. Now, we wish to find the best hyperplane which can separate the two classes. Data is not evenly distributed on left and right. Some of the triangle are on right too. You may feel we can ignore the two data points above 3rd hyperplane but that would be incorrect. SVM tries to find out maximum margin hyperplane but gives priority to correct classification.1st decision boundary is separating some triangle from circle but not all. It is not even

ISSN: 2233-7857 IJFGCN

showing good margin.2nd decision boundary is separating the data points similar to 1st boundary but here margin between boundary and data points is larger than the previous case.

Case 4: we will learn about outliers in SVM. We wish to find the best hyperplane which can separate the two classes. Data is not evenly distributed on left and right. Some of the triangle are on right too. In the real world, you may find few values that correspond to extreme cases i.e, exceptions. These exceptions are known as Outliers. SVM have the capability to detect and ignore outliers. In the image, 2 triangles are in between the group of circles. These triangles are outliers. While selecting hyperplane, SVM will automatically ignore these and select best-performing hyperplane.1st 2nd decision boundaries are separating classes but 1st decision boundary shows maximum margin in between boundary and support vectors.

VI. DATASET

The dataset is obtained from the Kaggle Data Science website called the 'Kaggle European Soccer Database'[12]. This database has been made publicly available and regroups data from three different sources, which have been scraped and collected in a usable database:

• Events, lineups & match scores : http://football-data.mx-api.enetscores.com/

• Betting chances: http://www.football-data.co.uk/

• Players and team features from EA Sports FIFA games: http://sofifa.com/

VII. CONCLUSION

The model we formulated depends on examination of Predicting result of soccer matches utilizing machine learning. We will have the option to make genuinely precise predictions. The accuracy of this model is quite acceptable by utilizing SVM algorithm. Thus, the came about Prediction System can investigates the utilization of machine learning techniques. The datasets gathered was effectively executed utilizing information mining procedure in various parts of the work. This exactness of framework can be additionally improved by including progressively significant highlights forming models which take into consider much more extensive parts of soccer.

ACKNOWLEDGEMENT

It gives us great pleasure in presenting the paper on 'Predicting outcome of soccer matches using machine learning'. I would like to take this opportunity to thank my internal guide Prof. A. B. Chandgude Department of Electronics and Telecommunication Engineering, SMT. KASHIBAI NAVALE COLLEGE OF ENGINEERING, for his unconditional guidance. I am really grateful to them for their kind support. We are thankful to Honorable Principal, SMT. KASHIBAI NAVALE COLLEGE OF ENGINEERING, Pune. Dr. A. V. Deshpande for the support. We are highly grateful to Dr. S. K. Jagtap Head of Department, Electronics and Telecommunication Engineering, SMT. KASHIBAI NAVALE COLLEGE OF ENGINEERING, for providing necessary facilities during the course of the work. We admit thanks to project coordinator, Department of Electronics and Telecommunication Engineering, for giving us such an opportunity to carry on such mind stimulating and innovative work

REFERENCES

[1] Maral Haghighat, Hamid Rastegari and Nasim Nourafza, "A Review of Data Mining Techniques for Result Prediction in Sports", ACSIJ Advances in Computer Science: An International Journal, Vol. 2, Issue 5, No.6, November 2013.

[2] Albina Yezus ,"Predicting outcome of soccer matches using machine learning", Saint Petersburg State University Mathematics and Mechanics Faculty,2014

[3] Brianne Boldrin ,"Predicting the Result of English Premier League Soccer Games with the use of Poisson Models", Mathematics and Computer Science of Stetson University,2017

[4] Norbert Danisik, Peter Lacko, Michal Farkas, "Football Match Prediction using Players Attributes", Online; accessed March 26th, 2018

[5] Siddhesh Sathe, Darshan Kasat, Neha Kulkarni, "Predictive Analysis of Premier League Using Machine Learning", International Journal of Innovative Research in Computer and Communication Engineering, 2017

[6] Ben Ulmer and Matthew Fernandez "Predicting Soccer Match Results in the English Premier League", 2017

[7] Tim van der Zaan, "Predicting the outcome of soccer matches in order to make money with betting", Business Analytics and Quantitative Marketing, Erasmus University Rotterdam, 2017

[8] Sumit Shrestha, "Premier League Game Result Prediction", Tribhuvan University, 2016

[9] Chinwe Peace and Nwachukwu, Enoch Okechukwu ,"An Improved Prediction System for Football a Match Result", IOSR Journal of Engineering (IOSR- JEN), Vol. 04, Issue 12 (December 2014)

[10] Albina Yezus ,"Predicting outcome of soccer matches using machine learning", Saint-Petersburg State University Mathematics and Mechanics Faculty , 2014

[11] Jongho Shin and Robert Gasparyan, "A novel way to Soccer Match Prediction", semantic scholar, 2014

[12] Maral Haghighat, Hamid Rastegari and Nasim Nourafza, "A Review of Data Mining Techniques for Result Prediction in Sports", ACSIJ Advances in Computer Science: an International Journal, Vol. 2, Issue 5, No.6 , November 2013

[13] Kaggle European Soccer Database [https://www.kaggle.com/hugomathien/ soccer]