International Journal of Research in Engineering, Science and Management
Volume-2, Issue-1, January-2019
www.ijresm.com | ISSN (Online): 2581-5792

Sports Result Prediction System

R. G. Sonkamble[1], T. A. Dasare[2], S. S. Patil[3], S. B. Patil[4], I. C. Pawar[5], A. M. Phadtare[6]
[1,2,3,4,5,6]*B.E. Student, Department of CSE, Sanjay Ghodawat Group of Institutions, Atigre, India*

***Abstract*: In the modern world, sports produces enough considerable statistical information about each player, team, games, and seasons. Earlier sports science was believed to be owned by experts, coaches, team managers, and analyzers. Sports organizations have recently realized that the availability of science in their data and want to take advantage of that science through the use of different data mining techniques. Sports data mining assists coaches and managers in different ways like prediction of the results, performance of the player, identifying the talent, and evaluation of game strategy. Prediction helps the managers and the clubs in making the right decision to win the leagues and the tournaments. The present study shows that earlier research on data mining systems to predict the results and evaluate the advantages and the disadvantages of each system. Prediction has been successfully applied in all sports. Although in many aspects, this application has been of very small limits. It is very important to look into the applications of the machine learning in these instances and see if its application can give better results in the analysis. By making use of datasets that is more precise and machine learning, this research aims to offer a solution that will help to make predictions be more accurate and precise than the earlier systems. Prediction is the heart of remarkable disciplines in science and that is the reason why philosophy of prediction is employed in many companies. Machine learning which an area of intelligent systems is will be used in this report to provide solutions to the problems in aspects of predicting the results of various sports.**

***Keywords*: Random Forest algorithm performing regression and Database.**

1. Introduction

Nowadays, the team managers and analysts give a prediction about who is going to win the match. In the present, different statistics are used for the prediction of football match result. The prediction of any football match is done by using some of the previous data. The prediction of the match depends on different variables like player stats, team stats; past data etc. which is used by club directors and managers to decide which team is going to win the match and what will be required to win the match. In the recent years, football result prediction has gained lots of popularity. Earlier, by gathering the features that affect the outcome of football matches, a predictive model Knowledge discovery in databases (KDD) was developed. Data mining techniques have also been used in the past. In existing systems, statistical and machine learning approaches were used in football match result prediction. For team management, the details of the team and the opponent team should be known by the manager and the other coaching staff members. The managers and the coaches should be able to identify strengths and weaknesses of different teams and accordingly prepare for the matches. The management should also be able take decisions during a match and to monitor a match in real time. For all these requirements, the prediction application should be able to provide ideas based upon the current situations. We have used multiple machine learning algorithms wherein Random Forest algorithm gave us the best accuracy. Random Forest Algorithm is used when the categorical variable is in different levels. Random forest algorithm is the most powerful supervised machine learning algorithm that is capable of performing regression and classification. The algorithm creates the forest with number of decision trees. The more the trees in the forest the more number of predictions and thus high accuracy. This method's main advantage is that it's very simple to explain the relationship between output variable and input. We can 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 English Premier League for the last one year and trained the algorithm based on that data. The algorithm learns from training data so that it forms different rules and pattern and based on that the algorithm makes a decision on some new data. The algorithm keeps on learning by using the feedback with every input. The software predicts the match result and starting 11 that is the lineup. The results of this software can be used by the management staff while team selection or by the fans playing while fantasy leagues in order to create an efficient team or for the purpose of betting.

2. Literature Survey

In the whole world football is the most famous and popular sport. Among all the sport, football prediction is the most widely researched. The main use of our project is to create software with which we can predict the upcoming match results based on many factors. We will give the teams name as the input and it will show which team will win the match. Not only who wins the match, the game lineup also will be also predicted. In the previous systems they used very less parameters which resulted in less accuracy of the prediction. We will be using many extra factors like player details, half-time score, full-time score and many more will be considered and the accuracy of the prediction increases. In the survey it was noticed that the use of SVM and KNN algorithm which gave in the range of 40% - 60% accuracy in match results using parameters like goals, corner kick and passes. Our aim is to bring the accuracy above 65%.

3. Methodology

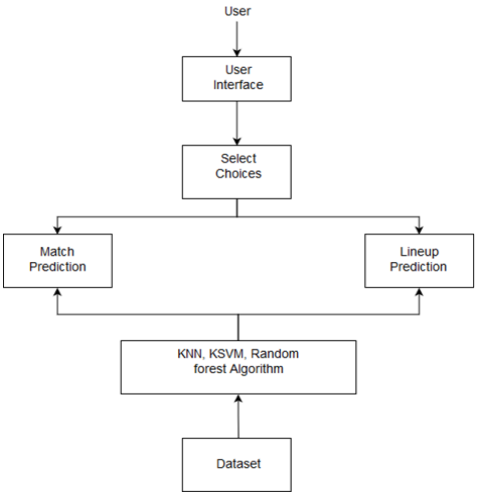

Fig. 1. Process flow Diagram

With the literature review, the most common thing was choosing the correct parameter is the first way to get the prediction. The more the parameters, the more chances of getting the result of the prediction correct. The prediction that was difficult for the experts have made some easy task due to several prediction methods. The parameters or factors such as home advantage, injuries of the players, cup game effect on league, team recent form and the head-to-head matches between the opponent need to be analyzed which adversely affect the result of the match. With lots of data, the data-mining tool is used to extract the information. Artificial Neural Network (ANN) and Regression techniques are two data-mining techniques that will be used. We are going to implement random forest algorithm. Random Forest Algorithm is used when the categorical variable is in different levels. Random forest algorithm is the most powerful supervised machine learning algorithm that is capable of performing regression and classification. The algorithm creates the forest with number of decision trees. The more the trees in the forest the more number of predictions and thus high accuracy. We will first collect the previous results of the matches with every history of the team and the teams they played with. Then, from the collected data will extract the features such as Home and Away Goal Difference, Points, Attack and Defense Skills, which is not needed for the techniques. Then a collective database will be built to collect all the necessary data and will be kept in MS Excel spreadsheet. So, with the known of the several factors, we came to use different input variables, output variables, algorithms for the process of prediction.

- Data collection and normalization
- Training the normalized data
- Testing the trained data
- Prediction

4. Conclusion

Considering the popularity of sports in the current world, many organization disburse large funds to gain better results in sports matches. Therefore, predicting the game results is now a subject of interest for different sports organizations. We implemented the model using different machine learning algorithms and were able to reach the accuracy of 80.6% with Random Forest Algorithm on the Match History Database of one year 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. We can also use the Machine Learning for the Goal Scorer prediction.

References

[1] Rory P. Bunker, Fadi Thabtah, “A machine learning framework for sport result prediction”, presented at Auckland University, Sept 2017,
[2] Ben Ulmer, Matthew Fernandez: “Predicting Soccer Match Results in the English Premier League”, Presented at Stanford University, California, Dec 2014,
[3] Data mining,
Online: https://en.wikipedia.org/wiki/Data_mining
[4] Machine Learning,
Online: https://en.wikipedia.org/wiki/Machine_learning.
[5] Lu WL, Ting JA, Little JJ, Murphy KP. “Learning to track and identify players from broadcast sports videos.” IEEE Transactions on Pattern Analysis and Machine Intelligence 2013
[6] Aditya Srinivas, and Timmaraju, A. P. (2013). “Game ON! Predicting English Premier League Match Outcomes.”
[7] Blundell, J. D. (2014). “Numerical Algorithms for Predicting Sports Results.” School of Computing, Faculty of Engineering.
[8] Yezus, A. (2014). “Predicting outcome of soccer matches using machine learning.” Saint-Petersburg State University, Mathematics and Mechanics Faculty.
[9] Dixon, M.J. and Coles, S.G. “Modelling association football scores and inefficiencies in the football betting market”,Journal of the Royal Statistical Society: Series C (Applied Statistics).
[10] Schumaker R. P, Solieman O. K, and Chen H. “Predictive modeling for sports and gaming.” 2010.