January-February 2020
ISSN: 0193-4120 Page No. 10630 -10633

A Novel Machine Learning Algorithm for Constructive Analysis of Cricket Data

[1]Sadiq Ali, [2]Rama. A, [3]N. Deepa

[1]UG Scholar, [2,3]Assistant Professor, Saveetha School of Engineering,
Saveetha Institute of Medical and technical Sciences, Chennai
[1]kingadiq193@gmail.com, [2]rama_j1@yahoo.com, [3]ndeepa.sse@saveetha.com

Article Info
Volume 82
Page Number: 10630 -10633
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 19 February 2020*

Abstract
Today from the youth to older peoples all are watching sports. Under the sports their different classification. But in India and most other countries all are concentrate in watching cricket. It is worldwide game played by different countries in various categories. Matches can be conducted as test match, T20, world cup etc. For each match the players can be selected based upon the past performance and the strike rating. The team is arranged by the captain he selects the 6 batsman and 5 bowlers. The size of the team is 11. The batsman can need to take maximum number of runs and the bowlers need to take maximum number of wickets and give minimum number of runs per over. This paper proposes the prediction of the wining team and prediction of the score of the batsman in both teams. This prediction can be made through the performance of the batsman in previous match and the performance of the bowler in previous match. Based upon the data analysis using SVM and random forest algorithm in machine learning technique the result can be determined.

Keywords: *Random forest, Support Vector Machine, Prediction, Navie Bayes.*

1. Introduction

Cricket is the sports which can turns concentration of all peoples towards it. Because it has one of the best professional game in which most of the countries gets contributes. In India nearly 92% of the people watching cricket peoples attracted toward it. The cricket matches can be organized by the cricket council team. The Indian cricket can make list of the members of around 20 players. From among 20 players the 11 members can be selected per team. The players are not remaining constant for each match based upon the performance in the past matches the players get varied. The players can be under the control of the captain of the cricket team. The team contains six batsman and five bowlers. The team should be good in both bating and bowling and wicket keeping. The main goal of the batsman during the matches is to score more runs in a minimum over and the main goal of the bowler is to take maximum number of wickets and to give less number of runs in each over. The batsman performance can be determined by the past performance, venue, pitch opponent team. The cricket game are classified as IPL, test match, league match, T20 etc. For each match number of rounds has been conducted and the winning in result is determined by the performance in each round. This paper shows the prediction of the winning team and the score of the batsman by the past performance.

The machine learning is a technique it plays a major in all the aspects. The machine learning is used to predict the result; it is one of the part of the artificial intelligence. The data of the previous match is get stored in the database. The data base has the individual score of each players of various team. By analyzing the data using the SVM, random forest, Bayers method. The result is get predicted and shown in the screen. These various algorithm can compares the data of the each players in the team with the past performance. The players are not remain constant for each matches based upon the performance in the past matches the players get varied. The players can be under the control of the captain of the cricket team. The variation is get stored in the particular path likewise each players data is get compared and stored in the path. For each match number of rounds has been conducted and the winning in result is determined by the performance in each round. This paper shows the prediction of the winning team and the score of the batsman by the past performance. The machine learning

January-February 2020
ISSN: 0193-4120 Page No. 10630 -10633

is a technique it plays a major in all the aspects. The machine learning is used to predict the result; it is one of the part of the artificial intelligence. The data of the previous match is get stored in the database. At the finalization of the overall result of the both team the wining result is get predicted. Among these various algorithms the random forest is the best and predicted the exact result.

2. Literature Survey

S. Muthuswamy et., al., proposed the data mining concept is get emerged in the field of the sports. For data analyst this technique has been implemented it can play a vital role in determining the wining percentage of the team before the match gets complete. But for the Bangladesh team match the result is not get predicted correctly. This prediction can be made in this paper by classification into three different stages before starting the match, after the first innings and the continuous wicket falling. They use final tree algorithm to determine the exact percentage of the result. This can shows the percentage in the three different stages in before starting the match the winning percentage is 80% and after the first innings the percentage is get reduced into 70% and after getting all wickets the percentage is get down into 35% at the final 2 over. This data analyst from the data base is predicted nearly 90% accurate by using the final tree method [1].

I. P. Wickramasinghe et., al., proposed the cricket matches most of the members can contributes some sponsorship to the each team and for the sessions. By sponsoring period these companies can makes their product advertise to the entire sector. By using the famous players and advertising their product it will reaches the most of the people and they preferred to by the product. In recent times Blue Chip Company is get sponsored for the redcap team. At the time these team are in peak the performance is good and their strike rate is also best compared to other players. Compared from the past ODI matches and local matches the performance is goes on increasing. The graph shows the random increase in points from the low range to high range. This can be monitored and stored in the database. In final match when the play against the Pakistan they defeat the opponent team and win the match. The prediction can be done using the data mining concept they use the previous data which is get stored in the server. From the past data they get compared with the present and predict the result of the winning team. The determined result of the redcap against the Pakistan is the nearly 87%. The prediction result can be absolutely close to the redcap team [2].

G. D. I. Barr and B. S. Kantor et., al., proposed most of the will waiting for the IPL match every time at the time of march to April. Each team has a separate owner they can inverse large amount of fund. They owners are leads into a pressure to choose the winning team. The wining of the team is depending upon the hands of the players in the team. So they want to select the players who are in the form. The owners can compitate with each other team and they select the players for their own teams. Selection of the players can finalize by the owner according certain criteria. But with all effort their judgment will goes wrong it will leads to great loss who get invest in the team. To overcome this here we use the data analyization technique to predict the winning team and players. The past data of the players get collected in the particular data path. The win, lose and the draw matches with each team is get noted in the path. The data can be compared with the current scenario and the players maximum winning percentage is get displayed with maximum runs they can score. The analyisation can be done using machine learning technology here the SVM algorithm is used which shows the maximum accurate result. The result shows to nearly 78% of the winning team [3].

S. R. Iyer and R. Sharda et., al., proposed most of the members watching the cricket in online nearly 20% of the whole are watching in live. At the time watching the peoples can predict the team which will get win best upon the batting performance and the wicket the prediction has been made. The predicted data is get stored in the cloud. The cloud is the big data server in which large amount of data gets gathered. By using the hadoop framework technique the data frame is get predicted. The data analyst can be made with reduce map system which can analyze the data with the peoples comments. The data is get distinguished into three divisions win, lose and draw. The data analyst cannot be done perfectly using the single algorithm the method is also used which is the trigraphics. It can collaborate with reduce map and founds a appropriate result with the most accurate correct result [4].

M. G. Jhanwar and V. Pudi et., al., proposed Before some years some of the matches are get fixed which has to win and lose. The players in the team and the official members of the team and umpire are get money from the opponent team and get lose in the match. They planned in the dressing room with all the internal and external members of the team. The players can enter into the stadium they wontedly beat the ball with high stamina for the normal ball and it get catch by the opponent team. These fixed matches can be predicted using the network which is the linking mining, graph mining, community system. They can noted the catches and runs of the team. By analyzing the data for the ball by ball they can predict that whether the match is fixed or the usual matches. To make the team to win the opponent team owner offered fund to the team who plays against their team and ask to lose the match. The fixing can be done in the final stage of the match whether it is semi final or final [5].

H. H. Lemmer et., al., proposed in current the social media and the news channels can plays a vital role. In social media they can post the opinion about the status of the match and the news channel can directly went to the people and collects the public opinion. By using the data in the social media we using the sentimental analysis the sentimental analysis can be done in English and various

January-February 2020
ISSN: 0193-4120 Page No. 10630 -10633

languages but it is get restricted it some stages in Bengali languages. So in this paper we propose that made a method and done sentimental analysis and get the data which is get classified into various categories. The opinion of the people in the Bangladesh to show the predicted data. The categories are the win, lose and the draw with the opponent team. The data gets collected in cloud by using some of the methods the unwanted data is get removed from the text. These process can be made through TDF vectorization method which can provides the overall opinion about the Bangladesh match [6].

D. Bhattacharjee and D. G. Pahinkar et., al., proposed the formulation of the table can be made through the k means algorithm and the clustering of the data between the year 2006-2012. The data in the specified period can be monitored. The performance of the player in each match against opponent team. The ball to ball run is stored in the particular path. The data in the path is get compared with the current performance of the player. The player in the team can determine the win or lose of the match. The performance can be predicted at each and every balls of the player. By using the data the table is get formulated which shows the current performance of the player and the winning result of the team. The data gets separated using k means algorithm method and for analyzing the data with the predicted result which can shows the accurate result of the team [7].

S. Mukherjee et., al., proposed the cricket games is famous worldwide but in Bangladesh is some more special. The people in the Bangladesh have the bigger fan of the cricket sports. The team can consist of batsman, bowler and wicketkeeper. The players in the team are selected for any international matches based upon the captain and the cricket squad. But the team will win or lose. In this paper they propose the making the winning squad by analyzing the last 2 years data of each players performance. Based upon the performance top 30 members has been selected. The performance of the player in each match against opponent team. The ball to ball run is stored in the particular path. The data in the path is get compared with the current performance of the player. The player in the team can determine the win or lose of the match. The genetic algorithm has been used to predict the winning team squad. From the 30 members the top 15 members has been selected as the final for the matches. The top selection has been created through the genetic algorithm. Each matches in the Bangladesh had follow this system to make the squad to win the final match [8].

P. Shah et., al., proposed the conversation are made through social media instead of face to face. During the matches the people can comment their prediction in their social media and natural way. By using the SVM, random forest and regression algorithm the winning prediction has been made. The data are collected from the website for every 10 over the tweets are monitored and clustered into one path and compare each comment to predict the result. Likewise for each set of period the tweets get monitored. The prediction is changed with regarding to time. For the last 10 over the data get closely monitored and predict the accurate result. The navies' bayes, regression algorithm plays a major role in predicting [9].

D. Parker, P. Burns et.,al., proposed most of the prediction method uses the machine learning algorithm and K means method. The grouping of the players in the team can be made through this method. The player's selection in the team can be done through the performance in the last few matches. The strike rate, score and the wicket taking based upon these criteria the players get selected. The selected players can be performed and the winning prediction can be made through the people opinion in the news and the social media based upon the opinion of the people and compare the result with one another based upon the result three section has been made one is the winning, losing and draw. The prediction result the percentage has been made to each part. The part which consists of higher percentage can be finalized as the result of prediction. The machine learning algorithm is plays a major role in various winning team prediction. By using this we made the team more stronger to defeat the opponent [10].

3. Materials and Methods

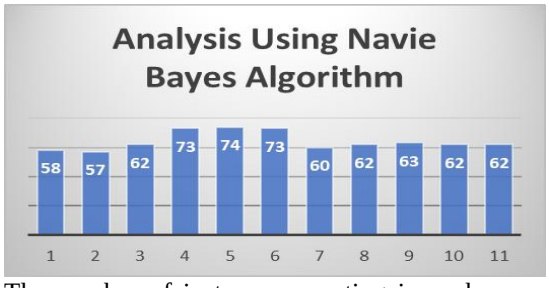
3.1 Dataset Description

id, season, city, date, team1 , toss_winner, toss_decision, result, dl_applied, winner, win_by_runs, win_by_wickets, player_of_match, venue.

3.2 Method Description

The data are analyzed data with respect to the city, climate, venue, scores etc.. During the international matches and IPL to predict the winning team data analyst has been made. The team is get arranged based upon the before performance of the players in various match. The data of each sportsman get stored in the database. By comparing the data with one another the sportsman gets selected. If the sportsman is get stronger the squad will be better. By using the random forest and navies bayes algorithm this data analysts can be made. Compared to various algorithm random forests is more better to predict the exact result.

4. Result and Discussion



The number of instances counting in each season, here presents the count number of instances in each winning team.

January-February 2020
ISSN: 0193-4120 Page No. 10630 -10633

5. Conclusion

This paper Analysis the winning criteria of the player depends on the toss winning of the corresponding game. Thus, it is demonstrating that the winning actually plays a great impact in winning the game statistically. There is relatively correlation in winning the toss and winning the game.

References

[1] S. Muthuswamy and S. S. Lam, "Bowler Performance Prediction for One-day International Cricket Using Neural Networks," in Industrial Engineering Research Conference, 2008.

[2] I. P. Wickramasinghe, "Predicting the performance of batsmen in test cricket," Journal of Human Sport & Excercise, vol. 9, no. 4, pp. 744-751, May 2014.

[3] G. D. I. Barr and B. S. Kantor, "A Criterion for Comparing and Selecting Batsmen in Limited Overs Cricket," Operational Research Society, vol. 55, no. 12, pp. 1266-1274, December 2004.

[4] S. R. Iyer and R. Sharda, "Prediction of athletes performance using neural networks: An application in cricket team selection," Expert Systems with Applications, vol. 36, pp. 5510-5522, April 2009.

[5] M. G. Jhanwar and V. Pudi, "Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach," in European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD 2016 2016), 2016.

[6] H. H. Lemmer, "The combined bowling rate as a measure of bowling performance in cricket," South African Journal for Research in Sport, Physical Education and Recreation, vol. 24, no. 2, pp. 37-44, January 2002.

[7] D. Bhattacharjee and D. G. Pahinkar, "Analysis of Performance of Bowlers using Combined Bowling Rate," International Journal of Sports Science and Engineering, vol. 6, no. 3, pp. 1750-9823, 2012.

[8] S. Mukherjee, "Quantifying individual performance in Cricket - A network analysis of batsmen and bowlers," Physica A: Statistical Mechanics and its Applications, vol. 393, pp. 624-637, 2014.

[9] P. Shah, "New performance measure in Cricket," ISOR Journal of Sports and Physical Education, vol. 4, no. 3, pp. 28-30, 2017.

[10] D. Parker, P. Burns and H. Natarajan, "Player valuations in the Indian Premier League," Frontier Economics, vol. 116, October 2008.