A SURVEY ON TEAM SELECTION IN GAME OF CRICKET USING MACHINE LEARNING

Vipul Punjabi[1], Rohit Chaudhari[2], Devendra Pal[3], Kunal Nhavi[4], Nikhil Shimpi[5], Harshal Joshi[6]

[1]*Assistant Professor, Dept. of Information Technology, R. C. Patel Institute of Technology, Shirpur (MS), India*

[2, 3, 4, 5, 6]*Students Dept. of Information Technology, R. C. Patel Institute of Technology, Shirpur (MS), India*

---***---

Abstract - *Player selection is one the most important tasks for any sport and cricket is no exception. The performance of the players depends on various factors such as the opposition team, the venue, his current form etc. The team management, the coach and the captain select 11 players for each match from a squad of 15 to 20 players. They analyses different characteristics and the statistics of the players to select the best playing 11 for each match. Each batsman contributes by scoring maximum runs possible and each bowler contributes by taking maximum wickets and minimum runs. This paper attempts to predict the performance of players as how many runs will each batsman score and how many wickets each bowler take for both the teams. Both the problems are targeted as classification problems where number of runs and number of wickets are classified in different ranges. We used Naive Bayes to generate the prediction models for both the problems.*

Key Words: **classification, performance,**

1. INTRODUCTION

Data mining is one of the widely used techniques for finding hidden patterns from voluminous data. Sports management committee uses data mining as a tool to select the players of the team to achieve best results. Cricket is considered today as one of the major world sports in terms of participants, spectators and media interest. Although it originates from England, cricket did not attract much interest and attention in Europe like football did. However it became hugely popular in countries such as India, Pakistan, Sri Lanka, Bangladesh, South Africa, Australia, New Zealand and West Indies, most of them former British colonies or still under the Crown influence. With an increased influence and interest in the game of cricket around the world, the International Cricket Conference (ICC) is trying to implement new development programs with the goal of producing more national teams capable of competing at Test level but also club teams that can compete in professional leagues at national or international level [1].

Thus, development of the shorter versions of the game such as the Twenty20 World Cup (2007), the official Indian Premier League (2008) and the Cricket Champion League (2009).Because of its increased popularity and tremendous developments, especially in terms of the birth of new professional competitions, cricket became today a major attraction, whose performance in all of its aspects is an important phenomenon to watch and measure. As a result, more applications and programs that monitor performance in cricket have already started to emerge.

Cricket is a sport played by two teams with each side having eleven players. Each team is a right blend of batsmen, ballers and all-rounder's. The batsmen's role is to score maximum runs possible and the ballers have to take maximum wickets and restrict the other team from scoring runs at the same time. All-rounder is the players who can both bat and ball and they contribute by scoring runs and taking wickets. Each player contributes towards the overall performance of the team by giving his best performance in each match. Each player's performance varies with factors like the team he is playing against and the ground at which the match is being played. It is important to select the right players that can perform the best in each match. The performance of a player also depends on several factors like his current form, his performance against a particular team, his performance at a particular venue etc. The team management, the coach and the captain analyse each player's characteristics, abilities and past stats to select the best 11 player for a given match.

In this paper, we predict the players' performance in matches by analysing their characteristics and stats using supervised machine learning techniques. For this, we predict batsmen's and ballers' performance separately as how many runs will a batsman score and how many wickets a baller will take in a particular match.

2. LITERATURE SURVEY

An extensive online search produced very few articles related to player's performance prediction in the game of cricket. A very small number of researchers have studied the performance of cricket players. Muthuswamy and Lam predicted the performance of Indian bowlers against seven international teams against which the Indian cricket team plays most frequently. They used back propagation network and radial basis network function to predict how many runs a bowler is likely to concede and how many wickets a bowler is likely to take in a given ODI match [2].

Wickramasinghe predicted the performance of batsmen in a test series using a hierarchical linear model. Using neural networks study of predicting "How many wickets will a bowler take?" can be possible, but their work was limited to only eight Indian bowlers and difficult to generalize for all the bowlers in the world. So some supervised machine learning algorithms to build prediction models that can be used to predict the performance of any player in a given match [3].

Iyer and Sharda[4] used neural networks to predict the performance of players where they classify batsmen and bowlers separately in three categories – performer, moderate and failure. Based on the number of times a player has received different ratings, they recommend if the player should be included in the team to play World Cup 2007. Jhanwar and Paudi[5] predict the outcome of a cricket match by comparing the strengths of the two teams. For this, they measured the performances of individual players of each team. They developed algorithms to model the performances of batsmen and bowlers where they determine the potential of a player by examining his career performance and then his recent performances. Lemmer[6] defined a new measure called Combined Bowling Rate to measure the performance of bowlers.

2.1 Batting and Balling Performance

To select the best combination of players for a team, it is necessary to develop a better selection tool. In the research of Kusiak, Kern, Kernstine and Tseng [7], they have applied K Nearest Neighbor to find a score which is closer to the standard score or threshold of a player. Mickey Arthur has summed up the power of participation or responsibility in a game victory by addressing some influencing factors of ODI cricket such as physical strength, scoring, home ground advantages, day/night situation, toss, and decision of batting first or second. How these factors will affect the performance of players in a cricket match is determined by understanding and estimation of the values used in Bayesian classifiers [8].Logistics and normal regression models suggested by Holder and Nevill to find out the ranking and home ground advantages for players as according to their research has found that players perform well when they play in their home grounds. Moreover, to determine the weak points of opposite teams, reverse data mining technique is used, which can plan defensive strategy for the next game the players will play [9]. The form and selection of the cricket players is defined in terms of average score, strike rate and consistent performance. A classification algorithm is used here which assigns classes to the best batsmen accordingly [10].

The combined bowling rate is a combination of three traditional bowling measures: bowling average, strike rate and economy. Bhattacharjee and Pahinkar [11] used this combined bowling rate to analyze the performance of bowlers in Indian Premier League (IPL). They also determined other factors that affect the performance of bowlers and applied multiple regression models to identify the factors that are empirically responsible for the performance of bowlers. Mukharjee[12]. Applied Social Network Analysis to rate batsmen and bowlers in a team performance. He generated a directed and weighted network of batsmen-bowlers using player-vs-player information available for test and ODI cricket. He also generated a network of batsmen and bowlers using the dismissal record of batsmen in the history of cricket. Shah [13] also defined new measures to measure players' performance. The new measure for batsmen takes into account the quality of each bowler he is facing and the new measure for bowlers considers the quality of each batsman he is bowling to.

A comprehensive review of the literature regarding the performance of both the player and the game reveals following findings. Stretch (2003) researched about cricketer's injuries that could affect their performances in the game. In this study, the author applied a hierarchical linear model (HLM) to model the nature of injuries to South African cricketers, including doctors and physiotherapists working with the South African team [14]. Kimber and Hansford (1993) proposed a method, which was based on nonparametric approach to assess the batting performance of cricket batsmen. Reaction time is regarded as one of the incalculable talents of cricketers in all the departments of the game [15]. Balasaheb, Maman, and Sandhu (2008) attempted to find the impact of visual skills training that could affect the performance of batsman. In their research, they showed how the visual skills improve the reaction time, depth perception, and eye of the cricketers, which eventually improves the batting performance of the player [16]. While discussing about necessary adjusted measures to analyze the player performance in the game of cricket, Lemmer (2008) investigated the performances of players' when players participate in small number of cricket matches.
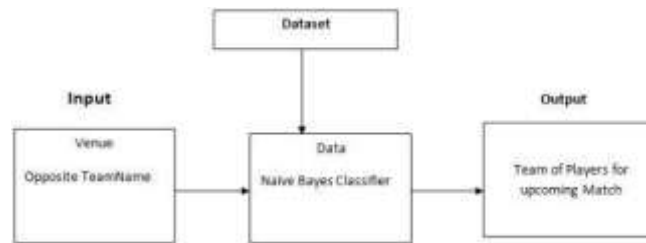
3. SYSTEM DEVELOPMENT

3.1 Proposed System

Fig -1: Proposed System

The system focuses on predicting the performance by analyzing their past performance. The input to the model is opposite team name and venue Decision Tree, Induction machine learning algorithm is used to generate the relevant attributes and also make the decision from the model. Data mining model of the proposed system is as depicted in above figure.

3.2 Data Preprossing

Calculating the Traditional Attributes:

The stats of the players such as average, strike rate etc. are not available directly for each game; we will calculate these attributes from the innings by innings list using aggregate functions and mathematical formulae. These attributes are generally used to measure a player's performance. These attributes are as follows:

3.2.1 Batting Attributes

a) **No of Innings:**

The number of innings in which the batsman has batted till the day of the match. This attribute signifies the experience of the batsman. The more innings the batsman has played, the more experienced the player is.

b) **Batting Average:**

Batting average commonly referred to as average is the average number of runs scored per innings. This attribute indicates the run scoring capability of the player.

Average = Runs Scored / Number of times dismissed

c) **Strike Rate (SR):**

Strike rate is the average number of runs scored per 100 balls faced. In limited overs cricket, it is important to score runs at a fast pace. More runs scored at a slow pace is rather harmful to the team as they have a limited number of overs. This attribute indicates how quickly the batsman can score runs.

Strike Rate: (Runs Scored / Balls Faced) * 100

d) **Centuries:**

Number of innings in which the batsman score more than 100 runs. This attribute indicates the capability of the player to play longer innings and score more runs.

e) **Fifties**:

Number of innings in which the batsman scored more than 50 (and less than 100) runs. This attribute indicates the capability of the player to play longer innings and score more runs.

f) **Zeros:**

Number of innings in which the batsman was dismissed without scoring a single run. This attributes shows how many times the batsman failed to score runs, hence this being a negative factor, was impacts the batsman's prediction negatively.

g) **Highest Score:**

The highest runs scored by a batsman in any (single) innings throughout his career. This attribute is used in the formula for calculating the venue attribute. This attribute shows the run scoring capability of the batsman at the venue. If a player has s very high score at a venue in past, he is more likely to score more runs at that venue.

3.2.2 Bowling Attributes

h) **No of Innings:**

The number of innings in which the bowler bowled at least one ball. It represents the bowling experience of a player. The more innings the player has played, the more experienced the player is.

i) **Overs:**

The number of overs bowled by a bowler. This attribute also indicates the experience of the bowler. The more overs the bowler has bowled, the more experienced the bowler is.

j) **Bowling Average**:

Bowling average is the number of runs conceded by a bowler per wicket taken. This attribute indicates the capabilities of the bowler to restrict the batsmen from scoring runs and taking wickets at the same time. Lower values of bowling average indicate more capabilities.

k) **Bowling Average:**

Number of runs conceded / Number of wickets taken.

l) **Bowling Strike Rate:**

Bowling strike rate is the number of balls bowled per wicket taken. This attribute indicates the wicket taking capability of the bowler. Lower values mean that the bowler is capable of taking wickets quickly.

m) **Strike Rate:**

Number of balls bowled / Number of wickets taken.

n) **Four/Five Wicket Haul:**

Number of innings in which the bowler has taken more than four wickets. This attribute indicates the capability of the bowler to take more wickets in an innings. Higher the value, more capable the player.

4. APPLICATIONS

1. Cricket team player selection in IPL where 1 team has to play different teams at different venue.

2. Player selection in World tournament such as ICC world cup or ICC champions Trophy where 1 team has to play different teams.

5. CONCLUSION

Selection of team especially in events where 1 team will be playing different teams in different conditions becomes very difficult for team management. Player's performance against different team varies as per venue. So keeping track of all this manually and then drawing a boundary line to select players is a very time consuming task. So the proposed system will help greatly in this aspect as it will provide entire player analysis with prediction of his performance on one click. Thus it will make the team selection process more accurate and precise.

6. FUTURE SCOPE

1. Currently we are using the IPL match dataset to show performance of the pro- posed methodology in future we can create different dataset to predict players for international level.

2. We can extend the system to manage dynamic data of matches where admin can add the data after each match so that the analysis is more dynamic.

REFERENCES

1. R. P. Schumaker, O. K. Solieman and H. Chen,"Predictive Modeling for Sports and Gaming," in Sports Data Mining, vol. 26, Boston, Massachusetts: Springer, 2010.

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

3. 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.

4. S. R. Iyer and R. Sharda, "Prediction of athletes performance using neural net- works: An application in cricket team selection," Expert Systems with Applica- tions, 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 (ECMLP- KDD 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. A. Kusiak, J. A. Kern, K. H. Kernstine, and B. T. L. Tseng, "Autonomous decision-making: A data mining approach," IEEE Transactions on Information Technology in Biomedicine, vol. 4, pp. 274–284, 2000.

8. "www.smh.com.au," Pietersen tells of racial discrimination - Cricket - Sport. [Online].Available:http://www.smh.com.au/news/ cricket/pietersen-tells ofracia   ld   is crimination/2006/08/30/1156816967716 .html. [Accessed: 10- Mar- 2018].

9. R. L. Holder and A. M. Nevill, "Modelling performance at international tennis and golf tournaments: is there a home advantage?" Journal of the Royal Statistical Society: Series D (The Statistician), vol. 46, no. 4, pp. 551–559, 1997.

10. H. H. Lemmer, "A measure for the batting performance of cricket players : research article," South African Journal for Research in Sport, Physical Education and Recreation, vol. 26, no. 1, 2004

11. 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.

12. 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.

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

14. Stretch, R. A. (2003). Cricket injuries: a longitudinal study of the nature of injuries to South African cricketers. British Journal of Sports Medicine, 37, pp.250-253

15. Kimber, A. C., Hansford, A. (1993). A Statistical Analysis of Batting in Cricket. Journal of the Royal Statistical Society, 156(3), pp.443-455.

16. Balasaheb, T., Maman, P., Sandhu J. S. (2008). The impact of visual skills training on batting performance in cricketers. Serbian Journal of Sports Sciences, 2(1), pp.17-23.