



CASE STUDY ON CRICKET TEAM PREDICTION

Gaurang Bagga¹

Affiliation:

- ^{1.} Jagan Institute of Management Studies Rohini Sector 5 Delhi-110085 India, gaurb1001@gmail.com

ABSTRACT

Cricket is a game of uncertainties. Love for the game of cricket is increasing day by day. In a country like India, it is more of an emotion. Some players are also considered as god. It's a game that is played between 2 teams of 11 players each into 3 different formats namely Test cricket—the purest form of the game and is also the longest format in all, and it is also this sport's biggest standard. A four innings match, which can last up to 5 days. It is generally referred as the testing of a team's Endurance and agility. One day match -It is a form of limited over match, in which both teams have a quota of 50 overs, team which scores the greatest number of runs wins the match (used to be 60 previously)with the game ending within a day If not obstructed by any environmental issues. T20 match -Twenty20 is the shortest form of the game usually completed within 4 hours. Both teams have a quota of 20 overs to score most numbers of runs to win the match. This paper focuses on finding a solution for cricket team prediction while analyzing team's data as per the conditions of the match that are home/away match, venues and toss decision. It relies on quantitative approach.

Key words: Cricket, prediction, Players, Match, Quantitative approach

1. INTRODUCTION

Cricket, a game of numbers - the runs scored by a batsman, the wicket taken by a bowler, the matches won by a particular cricket team, the number of times a batsman reacted in a certain way to a particular kind of bowling attack, etc. The ability to dig into this game's numbers for both improving performance and analyzing the other factors involved with this game. Cricket analytics provides interesting insights about the game and predictive intelligence regarding game outcomes.

Today, there are plenty of cricket game records and statistics available, e.g., ESPN cricinfo and cricbuzz. There are number of other such cricket databases that are used for cricket analysis using the newest machine learning and prognosticative modelling algorithms

Sports Analytics is a game changer when we talk about how actually these professional games are been played, especially about strategic decision, which previously was primarily done based on "gut feeling" or relativeness to past traditions. Sports Analytics is an enjoying field. Many researchers and companies find NumPy and other Py Data packages like Scikit-learn, SciPy, Matplotlib, and Jupyter easy for their purpose of working on the different models. Similarly, we have developed a model on r language using cricketr package, ggplot2 and dplyr.

Two types of analysis are- 1. Statistical Analysis: This helps in estimating the statistical significance of observational data in the context with various player and game tactics, estimating the game outcome by a generative or static model. Casual analysis and big data approaches are used for tactical analysis. 2. Data Visualization: Data graphing and visualization provides useful insights into relationship between various datasets. Previously, there was only one platform for cricket analytics that was provided by Clarke (1998). A lot of advancements have been done since that time. Advances in computing have provided more tools to exercise huge data sets. The structure of this article contains sections which are mostly independent and focuses on various topics of cricket. The paper concludes with some discussion on how cricket analytics may be heading

Agarwal, Yadav and Mehta (2017), suggested that the relative team strength between the competing teams forms a distinctive feature for predicting the winner. Modeling the team strength boils down to modeling individual player batting and bowling performances, forming the basis of approach used. Career statistics as well as the recent performances of a player have been used to model. Player **Ramakrishnan, Sethuraman, and Parameswaran (2019)** suggested that the relative team strength between the competing teams forms a distinctive feature for predicting the winner. Modeling the team strength boils down to modeling individual player's batting and bowling performances, forming the basis of our approach.

The case study to show data visualization for Sachin Tendulkar using r language using cricket package-

CODE:

```
1 install.packages("cricketr")
2 install.packages("sixer")
3 library("cricketr")
4 Tendulkar<- getPlayerData(35320,dir=".",
5 file="tendulkar.csv",type="batting",homeOrAway=c(1,2), result=c(1,2,4))
6 par(mfrow=c(1,3))
7 par(mar=c(4,4,2,2))
8 batsmanRunsFreqPerf("./tendulkar.csv","Sachin Tendulkar")
9 batsmanMeanStrikeRate("./tendulkar.csv","Sachin Tendulkar")
10 batsmanRunsRanges("./tendulkar.csv","Sachin Tendulkar")
11 battingPerf3d("./tendulkar.csv","Sachin Tendulkar")
12 batsmanAvgRunsGround("./tendulkar.csv","Sachin Tendulkar")
13 batsmanAvgRunsOpposition("./tendulkar.csv","Tendulkar")
```

Outputs

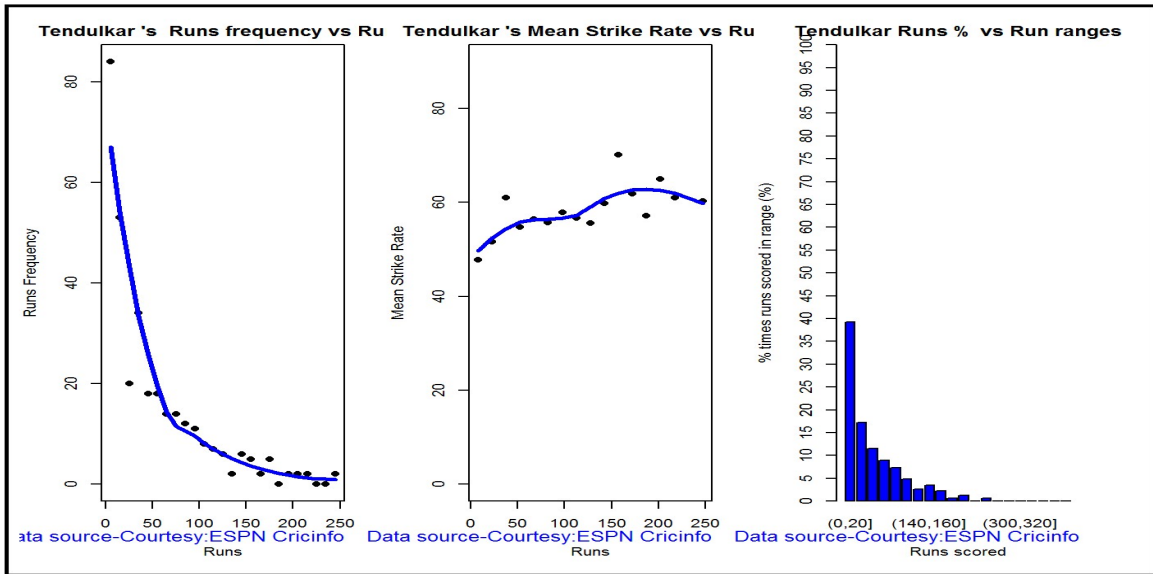


FIG 1.1 OUTPUT FOR LINE 8,9,10

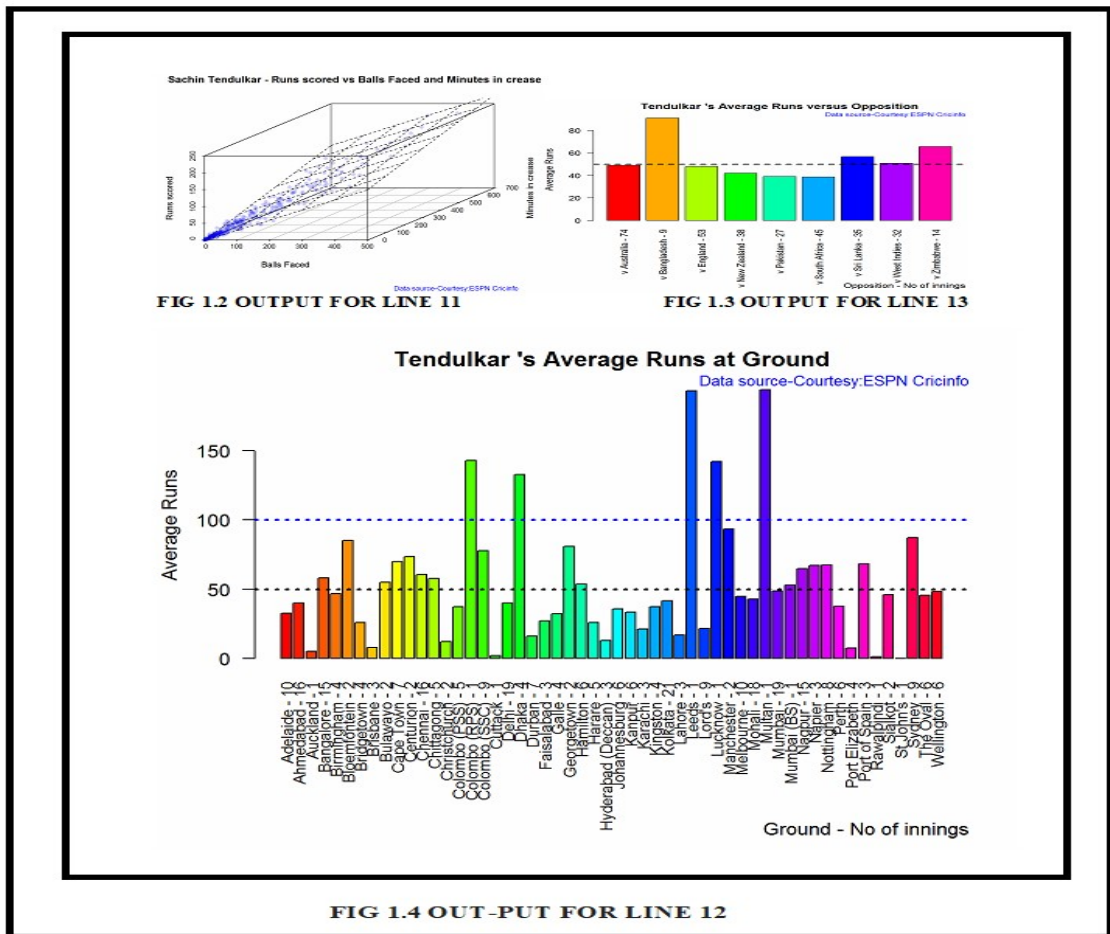


FIG 1.4 OUT-PUT FOR LINE 12

Nimagadda et al. (2018) applied statistical techniques to predict a T 20 match result while the match is in progress. a multiple regression model is developed to prediction purpose Using runs scored per over the innings, different algorithms such as Logistic Regression with multi-variable linear regression and Random Forest used to predict the final result. The software used here is Anaconda and Python libraries like pandas, NumPy and Python to work with the data structure and applying algorithms. The end result found was based on the impact of toss winner and match winner. The predictive model considered the innings score at regular period to found the pattern and predict accordingly. At the data found was satisfactory I.e., slightly above 50%.

Pathak & Wadhwa (2016) found the prediction of the result for matches using data mining techniques. They experimented on predicting the outcome for One Day International format based on various factors such as home ground, toss, innings, fitness of players and other strategies. In addition to the techniques implemented by SV Method was used to predict the result. they came with a tool called COP (Cricket Outcome Predictor), which provides the probability for winning an ODI match. The data understudy was done on all international cricket match played between 2001 to 2015. Results found clearly showed that the classifiers derived by the SVM method performed better than Naïve Bayes and Random Forests methods. SVM obtained accuracy, whereas the accuracy rates of the other methods were around 60%. The COP tool which is developed in R software helps a user to predict the match outcome and the user could analyze between the classifiers. A notable result was observed when COP system was applied on the India vs. Australia series in which Naïve Bayes derived more competitive classifiers in terms of predicting the match outcome.

Jhanwar & Pudi conducted a study to predict the for ODI cricket matches victimization data processing techniques. They investigated the match result victimization team players performance singly in batting and bowling aspects. Initially the potential of all players was understood victimization people career statistics and KNN, Support Vector Machine (SVM), Random Forests, logistical Regression and call trees techniques were applied. To predict the result of the match, the relative strength of every team is studied, in conjunction with the venue of the match and toss result. The accuracy of the KNN model was beyond the opposite models in predicting the game analytics for game results strength of the team players giving virtually 71 accuracy for the ODI match. There was no feature choice concerned during this study.

Kampakis & Thomas (2015) done a study for predicting the result of twenty over match format. The study was done on English Cricket Cup and also the model was tested on seasons between 2009 to 2014, supported the information from previous matches. A rule was designed on easy prediction and so any reports were administrated on features for in-depth analysis. foremost the team information was used and so player information was analyzed. choice ways used were Chi-square testing, mutual data and Pearson correlation. They used Naive Thomas Bayes, logistical Regression, Random Forests and Gradient call Trees on the chosen options from the information. By victimization these ways they foreseen the match outcome, it absolutely was found that the rule derived by Naïve Thomas Bayes offered around sixty fourth prediction accuracy on the dataset used. At a similar time comparing the accuracy of various techniques, Na €I ve Thomas Bayes made the very best level of accuracy, very cheap was Gradient call Trees.

Munir et al. (2015) experimented with twenty20 format to predict the result using data mining techniques. The main aim of this study is to the combine pre-game and in-game data for prediction. They considered the T 20 International match data along with IPL data until 2015 season for dataset. In depth analysis was conducted by putting the data on the basis of venue, one

team head-to-head results of teams, batting first and so on. Decision Tree was made to predict the match result, and it produced models with almost 78% accuracy for the team that bats first and 75% when it bats second. IG technique was used for feature selection.

2. METHODOLOGY

The work of our project focuses on 2 models.

2.1 Descriptive model

The descriptive model focuses principally on 2 aspects: It describes the information and statistics of the previous information concerning the batsmen, bowler or the all rounder. It provides the past data of the matches via by the groups.

2.2 Prognosticative model

The prognosticative model focuses on predicting the winning proportion of the team. The ranking of the players on the idea of points are displayed further. The user should select the 2 groups playing against one another. the choice of the teams works on the factors as:

- (i) If the players area unit batsmen, then, sorting is finished according to the strike rate of the batsmen.
- (ii) If the players area unit bowlers, then, sorting is finished according to the typical rate of the bowler.
- (iii) If the players area unit all rounder then, sorting is finished considering each strike rate further as average rate.

The rule utilized in this model is call Tree Classifier. a choice tree is made victimization top-down approach. during this rule the foundation node i.e., the prior issue thought-about is that the is that the match is being vie. The tree is made in step with the prominent factors (city, venue, teams, toss decision)considered within the match.

DECISION CLASSIFIER DIAGRAM:

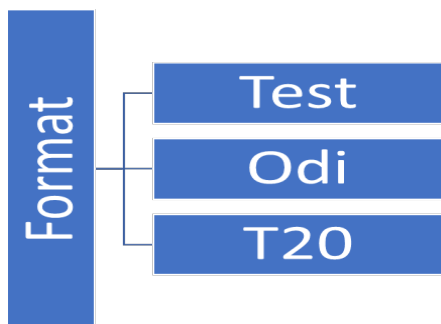


FIG 2.1 SELECTING FORMAT

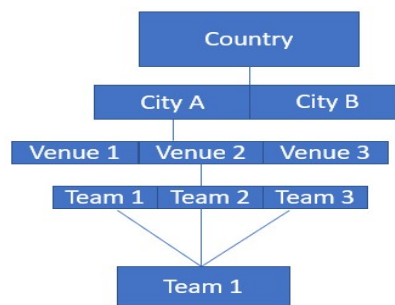


FIG 2.2 SELECTING CONDITIONS

Algorithm is as follows

- ▶ Start Select the format rest the algorithm goes same for everyone
- ▶ Select the node 'country'
- ▶ Choose from one of the cities from the option
- ▶ Select the venue of the match
- ▶ The teams are selected and compared with each other'
- ▶ End

3. RESULT

At the top the prediction is created in accordance with the descriptive approach and result given on the idea of a specific player and team's performance in numerous cities venues with the

result of toss because it tells that team can bat and bowl 1st. the user able to access the 2 models that's Descriptive model that shows the statistics of the player and also the prognosticative model that predicts the winning proportion of the team that the user has chosen.

4. CONCLUSIONS

This model will definitely assist the coaches and teams to select their squad as per the format and tournament. Not only helping the national teams but also for the domestic cricket and the famous leagues such as IPL, BBL and CPL etc. as it helps the stack holders so that they effectively choose their teams in auctions and bid on the players determining their capabilities and approach on the different venues under different conditions.

The most interesting part in this model has other aspects also that helps the normal public to make their teams to the platforms like dream11, game zy and my cricket circle etc. Similarly, it can help the bookies to determine their winning chances and helps them to invest their money but this is particularly not a good idea. This model is not intentionally made for this.

While identifying this model on the previous matches it is found that 60% of the teams supports this model and it is effectively working for the future matches also

5. FUTURE DIRECTION

In future the focus will be on adding the other concept that is sentiment analysis which will help in understanding what the public and experts wants to see. This will help further in understanding which team is best suited. This will cover the drawback of this model that is will help analyzing new players as there is no historical data available for them as they have no experience internationally or for the following format.

REFERENCES

- Agarwal, S., Yadav, L., and Mehta, S. (2017).** Cricket Team Prediction with Hadoop: applied math Modeling Approach. 5th International Conference on Information Technology and Quantitative Management, ITQM 2017 , Procedia Computer science, 122, 525–532.
- Analysis on attributes deciding cricket winning,** Retrieved from <https://www.scribd.com/document/357690109>
- Deep, C., Patvardhan, C., and Vasantha, C. (2016).** Data Analytics based mostly Deep dressing Predictor for IPL-9. International Journal of laptop Applications, 152 (6), 6–11.
- Duckworth, F. C. and Lewis, A. J. (2015).** A truthful technique for Resetting the Target in Interrupted One-Day Cricket Matches. Operational analysis Applied to Sports, 128–143.
- Jhanwar, M.G., Pudi, V. (2016).** 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).
- Kampakis, S., Thomas, W., (2015).** Using Machine Learning to Predict the Outcome of English County twenty over Cricket Matches. arXiv preprint arXiv:1511.05837.
- Munir, F., Hasan, M.K., Ahmed, S., Md Quraish, S., (2015).** Predicting a T-20 cricket match result while the match is in progress, Doctoral dissertation, BRAC University.
- Nimmagadda, A. Kalyan., Venkatesh, N.V. M., Teja, N.N.S., and Raju, C.G. (2018).** Cricket score and winning prediction using data mining. Int. J. Adv. Res. Development, 3 (3), 299–302.
- Passi, K. and Pandey, N. (2018).** Increased Prediction Accuracy within the Game of Cricket victimization Machine Learning. International Journal of knowledge Mining & information Management method, 8 (2), 19– 36.
- Pathak, N. and Wadhwa, H.(2016).** Applications of modern classification techniques to predict the outcome of ODI cricket, Procedia Comput. Science, 87, 55–60.

Predicting the Match Outcome in in some unspecified time in the future, Retrieved from <http://www.jssm.org/volume05/iss4/cap/jssm-05-480.pdf&p=DevEx.LB.1,5063.1>.

Ramakrishnan, Vijay., Sethuraman, K, and Parameswaran, R. (2019). Target Score Prediction within the game of Cricket. Retrived from://people.ucsc.edu/~praman1/static/pub/ML_Project_CS7641_report.pdf.

Satao, Preeti (2016). Cricket Score Prediction System (CSPS) victimization bunch algorithmic program. Technical analysis Organization Asian nation, 3(4), 2394– 0697.

Siddiqui, Tamanna., Alkadri, Mohammad., and Khan, Najeeb Ahmad (2017). Review of Programming Languages and Tools for giant for giant. International Journal of Advanced analysis in applied science, 8 (5), 8-10.

Singh, S. and Kaur, P. (2017). IPL image and Prediction victimization H-Base. Procedia applied science, 122, 910–915.