



CRICKET SCORE AND WINNING PREDICTION USING DATA MINING

Kushoo¹, Nisha², Shilpa Abhang³

¹Student, Department of MCA, Jyoti Nivas College, Bangalore, India

²Student, Department of MCA, Jyoti Nivas College, Bangalore, India

³Faculty, Department of MCA, Jyoti Nivas College, Bangalore, India

ABSTRACT

Cricket is one of the most popular team games in the world. Cricket is the second most watched sport in the world after soccer, and enjoys a multi-million dollar industry. There is striking enthusiasm for reenacting cricket and all the more critically in foreseeing the result of cricket coordinate which is played in three organizations to be specific test coordinate, one day universal and T20 coordinate. The perplexing standards winning in the game, alongside the different normal parameters influencing the result of a cricket coordinate present huge difficulties for exact expectation.

I. INTRODUCTION

Cricket is being played in numerous nations all around the globe. There are a ton of local and global competitions being held in numerous nations which play cricket. Cricket is a game played between two groups containing 11 players in each group. The outcome is either a success, misfortune or a tie.

In any case, at times because of terrible climate conditions the game is additionally cleaned out as Cricket is a game which can't be played in downpour. In addition, this game is likewise very eccentric on the grounds that at each phase of the game the force movements to one of the groups between the two. A great deal of times the outcome gets this show on the road on the last bundle of the match where the game gets truly close. Thinking about all these flighty situations of this unusual game, there is a gigantic enthusiasm among the onlookers to do some expectation either toward the beginning of the game or during the game. Numerous observers additionally play wagering matches to dominate cash. Along these lines, remembering every one of these potential outcomes, this report targets concentrating the issue of anticipating the game outcomes before the game has begun dependent on the insights and information accessible from the informational collection

The information mining calculations which can be applied on those properties. The game forecast issue that I am contemplating doesn't mull over the player's presentation yet it takes into thought the group's past exhibition at an elevated level degree alongside different variables like hurl champ, hurl choice, home help, and so on. The information mining calculations that are utilized are Decision Tree, Random Forest, Naïve Bayes, K-Nearest Neighbor.

II. LITERATURE REVIEW

The information mining calculations which can be applied on those properties. The game expectation issue that I am contemplating doesn't mull over the player's performance but it does take into consideration the team's past performance at a high level extent along with the other factors like toss winner, toss decision, home support, etc.

In paper [1] authors Found, three classification algorithms and compared to find the best accurate algorithm. The implementation tools used are Anaconda navigator and Jupyter. Random Forest is observed to be the best accurate classifier with 89.15% to predict the best player performance. Hence using this prediction, the best team can be formed.

In paper [2] authors used par score concept given by Duckworth & Lewis. This probability considers balls faced, balls left, runs scored, runs left, wicket, wickets left.

In paper [3] authors Presented a model to predict the outcome of an ODI cricket match while the game is in progress. It used the data of previous matches played between the team in order to design our model and Multiple Variable Linear Regression to design this model.

In paper [4], authors seen that winning the coin toss gives competitive advantage for Day & Night matches. It was found that SVM was proved to be a better model based on both the parameters used- accuracy and model outcome.

In paper [5], authors present model for predicting the final score of the first innings and estimating the outcome of the match in the second innings for the limited overs cricket match. Factors like the toss, the ODI ranking of the teams and the home team advantage will be considered in the predictions.

In paper [6] authors used win predictions techniques. This method contains dynamic team properties for win prediction like player history, winning percentage or ground history as well. We evaluated this technique over 100 matches as results are very interesting because of 85% correct predictions.

In paper [7] authors used 12 factors for each team for predicting the outcome of a game. Also, it has been observed that the home and away conditions greatly influence the outcome of the game. Hence the factors related to home and away conditions have been assigned greater weights. That is a prediction accuracy of exactly 70%.

In paper [8] authors used four attributes the name of the home team, name of the opponent team, match results, toss results. It used different supervised learning algorithms such as Logistic regression, support vector machine, Decision tree, and Bayes point machine binary classification model.

In paper [9] authors used two method predicts the score of first innings based on current run rate and also considers number of wickets fallen, batting team and venue of the match. The accuracy of the Naïve Bayes for predicting the outcome of the match, goes from 70% (initially) to 91% as the match progresses.

In paper [10] authors used attributes like Pitch, Team Strength, Weather, Venue etc. and in-game attributed like run rate, total run, strike rate, wickets in hand etc. influence a match result predominantly

In paper [11] authors concluded that Teams should effectively consume data to derive insights. Predicting the entire outcome of the game is difficult, but technology can assist in doing so. Human decision making still dominates technology as most teams have similar data but performance varies.

In paper [12] authors used factor like weather and the location. i.e. the country and the stadium where Duckworth Lewis method had come to use. The dataset consists mainly of One Day International (ODI) matches of teams from India, Pakistan, England, West Indies, South Africa, New Zealand, Sri Lanka, Australia, Bangladesh & Afghanistan. This paper presents a novel approach to evaluate the Duckworth Lewis system which is used to predict the target score in rain affected cricket matches when one or both the teams have had their innings shortened.

Table 1 depicts the existing techniques used

Table1: Existing Methods

Paper	Techniques Used
[1]D. Jyothsna, K. Srikanth, Analysing and Predicting outcome of IPL Cricket Data	Classification algorithm, K-means clustering Decision Tree, Linear Regression.
[2] Parag Shah, Predicting Outcome of Live Cricket Match Using Duckworth-Lewis Par Score	K-means clustering, Classification algorithm, Linear Regression
[3] Nimmagadda Akhil et.al, Cricket score and winning prediction using data mining	K-means clustering algorithm Logical regression Decision Tree
[4] Rameshwari Lokhande and Pramila M. Chawan, Live Cricket Score and Winning Prediction	Decision Tree ,Linear Regression, Pattern association, Pattern classification Decision Tree
[5] 6: Muhammad Yasir, LI CHEN, Sabir Ali Shah, Khalid Akbar, and M.Umer Sarwar, Match Prediction in T20 International	Linear Regression, K-means clustering, Classification algorithm.

III. PROPOSED METHODOLOGY

In sports, most of the prediction job is done using regression or classification tasks, both of which come under supervised learning. In Regression, the output is a continuous value; however, classification deals with discrete kind of output. For predicting continuous values, Linear Regression appeared to be quite effective, and for classification problems like predicting the outcome of matches or classifying players, learning algorithms like Naive Bayes, Logistic Regression, Neural Networks, and Random Forests were found being used in most of the previous studies. In this work, the various factors that affect the outcome of a cricket match were analyzed, and it was observed that home team, away team, venue, toss winner, toss decision, home team weight, away team weight, influence the win probability of

a team. The proposed prediction model makes use of multivariate Regression to calculate points of each player in the league and compute the overall strength of each team based on the past performance of the players who have appeared most for the team.

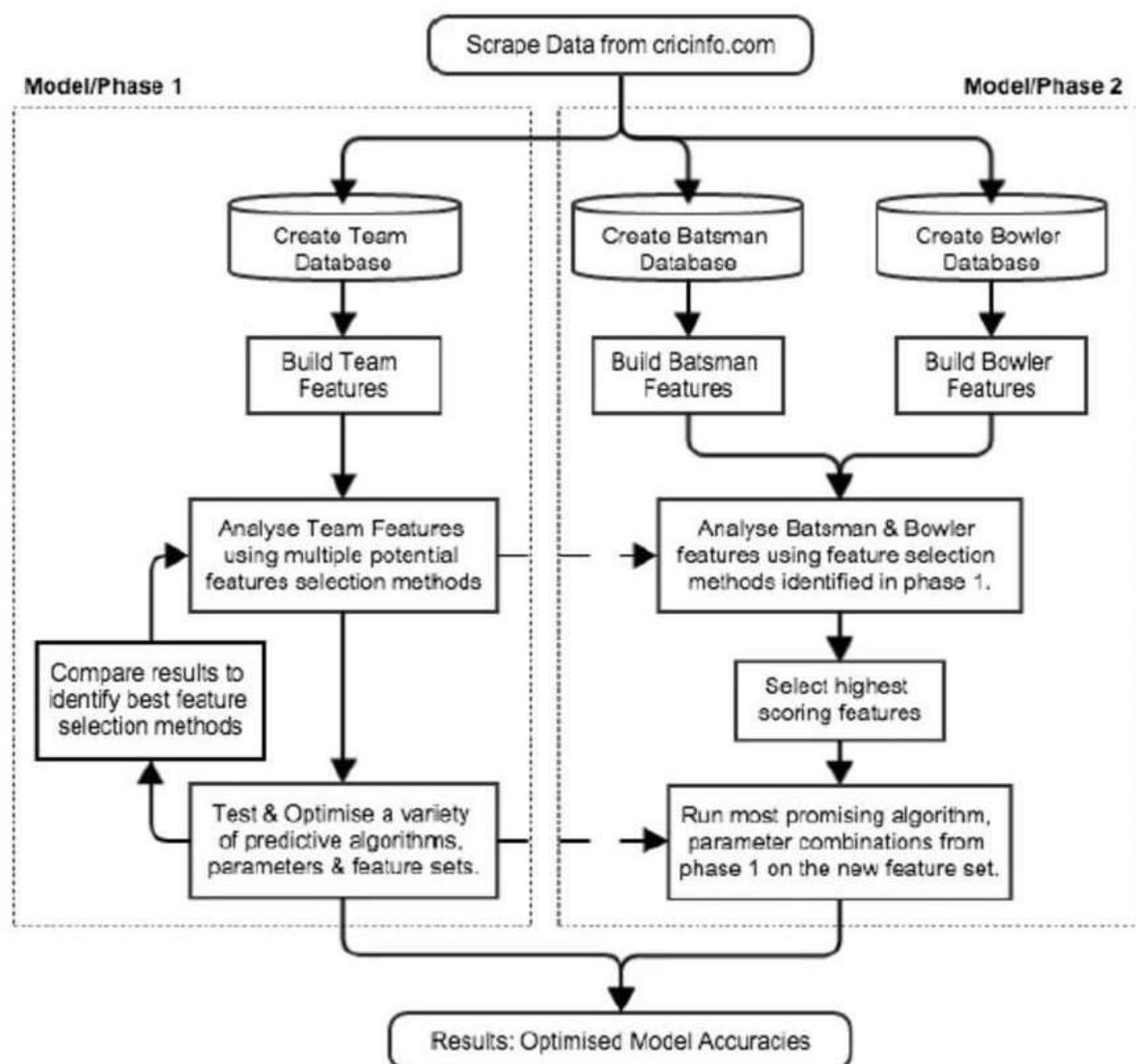


Figure 1. Methodology flowchart

IV. IMPLEMENTATION AND RESULTS

Environmental Setup: Python SciPy environment installed ideally with Python 3.6.3, is installed, and other libraries like Scikit-learn, Pandas, NumPy, and Matplotlib is installed. Minimum of 8GB RAM is required. VGG16 Algorithm is used for feature extraction

Dataset:

- The Cricsheet website is open source.
- We stored the data in separate csv files.

- We imported all the libraries and dependencies.
- We have loaded the csv file containing the details of each team's history in previous matches.
- We also loaded the csv file containing the results of matches played.
- This was done as the results of the last few years should only matter for our predictions.
- Then we did manual cleaning of the data as per my needs to make a machine learning model out of it.

1	JA Morkel9587021.36467236467	1	A Ashish Reddy34783992322218
2	T Thushara12190.631578947368	2	A Chandila242105911011710
3	V Sehwag272818331.48827059465	3	A Flintoff10521274077
4	Shivam Sharma531.66666666667	4	A Kumble10274104065737235
5	NK Patel1211240.975806451613	5	A Mishra25208519041624143118
6	GB Hogg22221.0	6	A Mithun4481141171714221
7	GD McGrath450.8	7	A Nehra203875753491522865
8	CK Kapugedera16240.666666666667	8	A Nel31751032
9	DE Bollinger21230.913043478261	9	A Singh6032121484116717
10	A Singh2100.2	10	A Symonds6562192073614433
11	CRD Fernando431.33333333333	11	A Uniyal6411144063
12	M Kartik1131091.03669724771	12	AA Chavan318117791412816
13	RE Levi83741.12162162162	13	AA Jhunjunwala1263131100105
14	MJ Lumb2781981.40404040404	14	AA Kazi20362011
15	Iqbal Abdulla46520.884615384615	15	AA Noffke358112022
16	SK Warne1982190.904109589041	16	AB Agarkar111231030546411341
17	JDS Neesham42460.913043478261	17	AB Dinda165158942470619347
18	MC Henriques5104021.26865671642	18	AB McDonald25954931002011
19	S Nadeem23570.40350877193	19	Abdur Razzak23631031
20	Kamran Khan350.6	20	AC Gilchrist0100000
21	KW Richardson35380.921052631579	21	AC Thomas3941431042614111
22	PC Valthaty5054321.16898148148	22	AC Voges7417281082

RESULTS

PREDICTED :

Target : 176
Score : 123

Winner :Team 1

Team 1

AM Nayar - 40
RV Uthappa - 39
DJ Bravo - 23

Team 2

Y Venugopal Rao - 44
DB Ravi Teja - 26
AC Gilchrist - 19

ACTUAL :

Team 1 178
Team 2 153

Innings 1 AM Nayar - 37
Innings 1 Y Venugopal Rao - 57

V. CONCLUSION

Selection of the right players for each match plays a significant role in a team's victory. An accurate prediction of how many runs a batsman is likely to score and how many wickets a bowler is likely to take in a match will help the team management select best players for each match. We modelled batting and bowling datasets based on players' stats and Characteristics. Random Forest turned out to be the most accurate classifier for both the datasets with an accuracy of 90.74% for predicting runs scored by a batsman and 92.25% for predicting wickets taken by a bowler. Results of SVM were surprising as it achieved an accuracy of just 51.45% for predicting runs and 70.95% for predicting wickets.

REFERENCES

1. Dataset: <http://cricsheet.org> [Online accessed 02-September-2016]
2. Duckworth, Frank C., and Anthony J. Lewis. "A fair method for resetting the target in interrupted one-day cricket matches." *Journal of the Operational Research Society* 49.3 (1998): 220-227.
3. Beaudoin, David, and Tim B. Swartz. "The best batsmen and bowlers in one-day cricket." *South African Statistical Journal* 37.2 (2003): 203.
4. Lewis, A. J. "Towards fairer measures of player performance in one-day cricket." *Journal of the Operational Research Society* 56.7 (2005): 804-815.
5. Swartz, Tim B., Paramjit S. Gill, and David Beaudoin. "Optimal batting orders in one-day cricket." *Computers and operations research* 33.7 (2006): 1939-1950.
6. Norman, John M., and Stephen R. Clarke. "Optimal batting orders in cricket." *Journal of the Operational Research Society* 61.6 (2010): 980-986.
7. Kimber, Alan. "A graphical display for comparing bowlers in cricket." *Teaching Statistics* 15.3 (1993): 84-86.
8. Barr, G. D. I., and B. S. Kantor. "A criterion for comparing and selecting batsmen in limited overs cricket." *Journal of the Operational Research Society* 55.12 (2004): 1266-1274.
9. Van Staden, Paul Jacobus. "Comparison of cricketers bowling and batting performances using graphical displays." (2009)
10. Trawinski, Krzysztof. "A fuzzy classification system for prediction of the results of the basketball games." *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*. IEEE, 2010.
11. Haghghat, Maral, Hamid Rastegari, and Nasim Nourafza. "A review of data mining techniques for result prediction in sports." *Advances in Computer Science: an International Journal* 2.5 (2013): 7-12.
12. Zdravevski, Eftim, and Andrea Kulakov. "System for Prediction of the Winner in a Sports Game." *ICT Innovations 2009*. Springer Berlin Heidelberg, 2010. 55-63.
13. <http://espncricinfo.com/> [Online accessed 20-September-2016]
14. Kaluarachchi, Amal, and S. Varde Aparna. "CricAI: A classification based tool to predict the outcome in ODI cricket." *2010 Fifth International Conference on Information and Automation for Sustainability*. IEEE, 2010.
15. Sankaranarayanan, Vignesh Veppur, Junaed Sattar, and Laks VS Lakshmanan. "Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction." *SDM*. 2014.