

Cricket Score Prediction Using Machine Learning

Prof. R. R. Kamble¹, Nidhi Koul², Kaustubh Adhav³, Akshay Dixit⁴, Rutuja Pakhare⁵

¹RSCOE, Assistant Professor, Pune, Maharashtra, India.

^{2,3,4,5}RSCOE, Pune, Maharashtra, India.

Article History: Received: 10 November 2020; Revised: 12 January 2021; Accepted: 27 January 2021; Published online: 05 April 2021

Abstract Currently, there is a system which can calculate the current run rate and from it calculates the final score of the team. It doesn't consider the fact about the no of wickets and also where the game is being played. The problem with the current system is that it is unable to predict the score of the 2nd team and also unable to predict the win percentage This system which is developed will have 2 model in it the 1st model predict the score a team will get after playing 50 over from the current situation. The second method predicts the win percentage of both teams even before the match has started this done by player selection. We found that error in regression toward the mean classifier could be a smaller quantity than Naïve mathematician in predicting match outcome has been sixty-eight ab initio from 2-15 overs to ninety-one until the top of 42th over.

Keywords: Cricket Prediction, Cricket analysis, Linear Regression, Naïve Bayes.

1. Introduction

Cricket is the foremost loved game, after football. The sport begins in England within the sixteenth century. Today, it's not just a game in India it's a religion with more and more fan increasing the world the game cricket may even overtake football as the most fan base sport in the world. There are 3 format namely. An ODI match which consist of 50 over which is completed in a day. Second being the test format this is the old format of the game it is played in 5 days which consist of 2 inning from both team it consist of 80 to 90 over game in a day. The team has to consistently perform for 5 days. This is very challenging format of the game where a player endurance, strength, patience and mentality matter the most. The 3rd and the newest format of cricket is the t20 format. This format was introduced in the 2006 and had its first world cup in 2007 which was won by India. It is a fast game which consist of 20 over this game is over under 3 hour. It consist of 2 team where each team get 20 over to play. The t20 format is very popular in India because of IPL. This tournament is the reason for the rise of t20 format in India.

This system which is developed will have 2 model in it the 1st model predict the score a team will get after playing 50 over from the current situation. The second method predicts the win percentage of both teams even before the match has started this done by player selection. The elemental objective of this paper is to consolidate past matches data and in-diversion data keeping in mind the top goal to arrange the simplest discerning model. For outlining the nice discerning model: range of wickets fallen, setting of the match, positioning of the batting cluster, hurl, and home group advantage these parameters are thought of. As no such analysis has improved true this organization of cricket, thus this may be chosen to require the take a look at as cricket matches are particularly distinguished these days.

2. Related Work

A. Predicting the Outcome of ODI Cricket Matches:

While team composition is predominantly based on the application of mathematical modeling in sports, predicting the outcome of a game is a fundamental flaw. Cricket is one of the most comprehensive team games on the planet's break. With this lesson, we have a tendency to start predicting one-day international (ODI) match results using a supervised learning approach from a team composition perspective. We have a tendency to use career statistics because player performance has modeled him recently. Players predict the outcome of the match along with the freelance factors

B. Score and Winning Prediction in Cricket through Data Mining:

Currently, the initial innings score supports a predetermination of an international (ODI) match for some purpose. The current run rate can be calculated, resulting in the number of runs bowled. It does not take into account the amount of wickets dropped and the venue for the match. There is no method of

predicting match results between the second innings. A model is introduced throughout this paper, which includes 2 methods, although the initial innings score does not fully support the current run rate, although it does take into account the number of wickets, the position of the match and the batting team. Is. The second method ts the outcome of the match between the second innings, a goal given to the batting team taking into account the same characteristics as the previous one. These two methods have been applied to the initial shift and the second inning to the worst-case linear regression classification and the Novav Thomas Bayes classification. In each method, fifteen overs breaks are scored in the fifty overs of the match, and the most mentioned features are recorded at each interval, which are recorded in all non-court matches between 2004 and 2014.

C. Player’s Performance Prediction in ODI Cricket Using Machine Learning Algorithms

This paper presents a simplified methodology for assessing the performance of a cricket player's upcoming match by applying machine learning algorithms. The job model includes applied mathematical knowledge of Indian national cricket team players sourced from support vector machines with reliable sports websites, algorithm feature elimination and linear vector regression, linear and polynomial kernels. To implement the planned model, the applied mathematical knowledge collected is processed as a numerical value to be implemented between the algorithms. Furthermore, the aforementioned Feature Choice Algorithm applies the field unit to extract properties related to the Area Unit Output feature. In addition, the machine learning algorithm uses the field unit to predict the runs scored by a hitter and the bowler's thought between the coming match. The experimental setup model provides ninety-one-half% accuracy for the hitter Virat and seventy-five% for the bowler Jadeja.

3. Proposed System

In this system we are going to predict the outcome as well as the score after 50 over from the current situation. Also we are going to predict the win percentage of the teams who are playing the game based on the squad selected. This system will not be only beneficial for the audience but also for the team selector as they can predict the win percentage on a particular team.

Advantage:

1. Predict the cricket score
2. Effective prediction technique
3. Predict the win percentage.

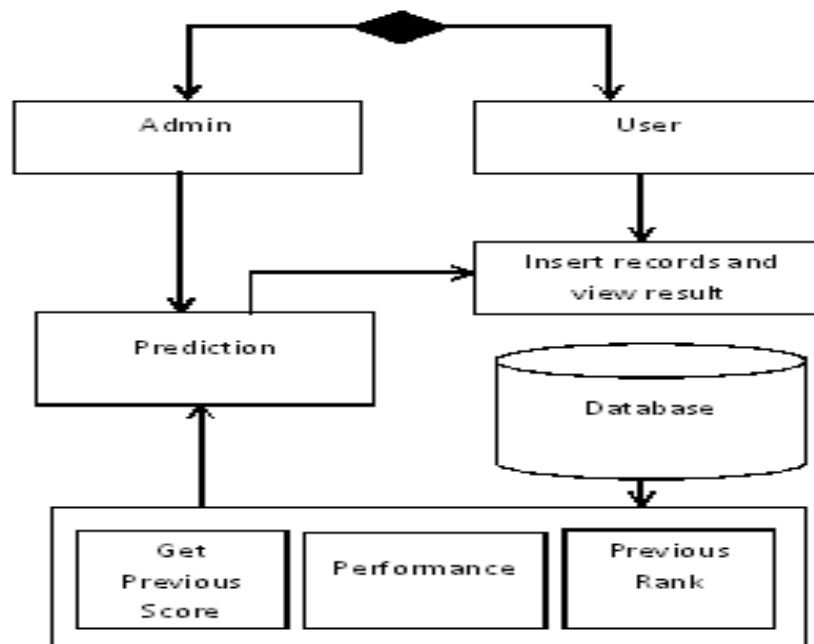


Figure 1. Proposed System

Winning the Toss:

The importance of this point is that when a captain win a toss the decision he will take whether to bat or bowl will affect the score and win prediction. This is because some ground are batting ground and some are bowling ground. This will affect the outcome of the game.

Ground:

The ground where the game is played is also important as the game may be home game or a away game or it may be a game played in some other ground where both the team don't belong this impact the player performance and eventually the outcome of the game. This point will help us to analyze the player mentality as well as their performance in home game or a away game.

This system tries to predict the score and winning chance a team has over another team. we tend to use Naïve Thomas Bayes, Random Forest, multiclass SVM, and call Tree classifiers to induce the prediction models for each the problems. Random Forest classifier was found to be the foremost correct for each issue. It'll facilitate predict the score of the team from the present over to fiftieth over in AN ODI match. It'll use the present over, current runs scored, no of wickets, runs created by the striker, runs created by the non-striker.

This system also will facilitate in predicting the win share of the 2 teams. For this, we've got used information from the last seventeen years (2000-2017). This information is taken from the espn.com with the assistance of net scraping. This we tend to change the United States to match each the groups on the concept of their player performance that embrace run rate, strike rate, no of runs, last season performance for the batsmen. For the bowler and every one rounder we'll be ready to compare on the idea of wickets taken, economy, what percentage ball needed to require a wicket. For score prediction data is taken from the info base which contains the last 5 year performance of all teams and their team member. When the user enter the no of runs, over, wicket and press the calculate button it'll immediately process the info given by the user and begin calculating on the idea of the database this may enable for a accurate prediction of score.

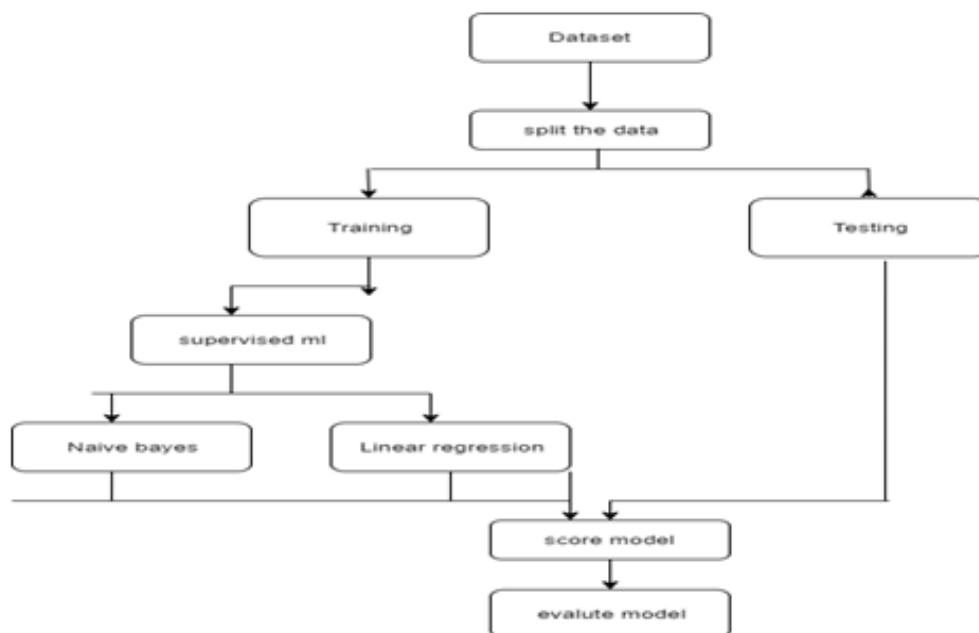


Figure 2. Flow Chart Diagram

A. Dataset

Consist of data of last 5 year for score prediction. For win prediction the data is for 17 year

B. Split the data

Divide the data into training and testing part we have divide in 70 -30 ratio .70 percent for training and 30 percent for testing of the data.

C. Training

Training is done using training data which is received from dataset. this training will help the system to learn from the data about the pattern and various relationship.

D. Testing

Testing of data is done to test whether the training phase has been successful or not the testing data is used to test the data after the training this ensure that the prediction or calculation done by the machine learning algorithm is right or wrong .

E. Supervised machine learning

It is used to train the algorithm to perform same task on various data to extract various pattern and relationship from them. Supervised learning provide data with eg and result to train the algorithm.

F. Naïve bayes

They are extremely ascendable, requiring a sort of parameters linear at intervals the number of variables throughout a learning downside. Maximum-likelihood coaching square measure typically done by evaluating a closed-form expression, that takes linear time, rather than by high-priced unvarying approximation as used for many alternative varieties of classifiers.

G. Linear regression

It repeatedly keeps performing the same task to provide with better prediction. This model is used for prediction and calculating the future values.

H. Score model

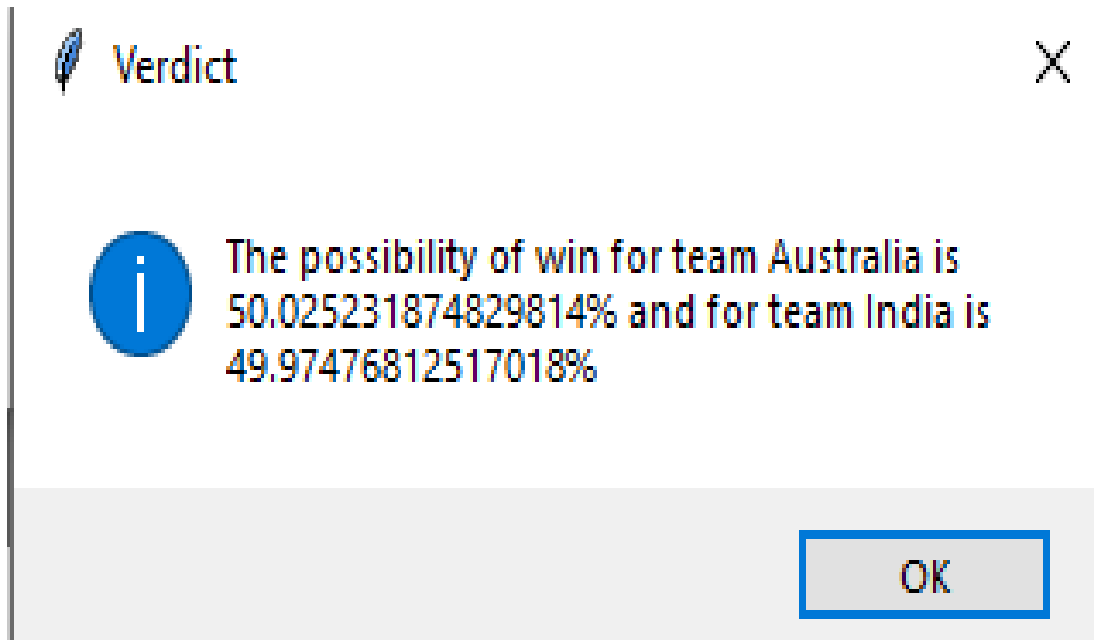
This model shows the numeric value after calculation and prediction made by various algorithm.

I. Evaluate model

This model is used to evaluate whether the prediction is right or wrong.

4. Results

The screenshot shows a web application window titled "Win Loss Prediction". The main heading is "Give Details To Scrap". Below this, there are two dropdown menus: "Australia" and "India", separated by "Vs". Underneath, there are three date pickers: "17", "Jan", and "2017". Below the date pickers, there is a text input field labeled "Enter the ground ID of the Match" with the value "292" entered. At the bottom, there is a "Next" button.



Cricket Score Prediction

| |
|----------------|
| 292 |
| 2 |
| 30 |
| 18 |
| 30 |
| PREDICT |

cricket score predict should be \$ 442.537

Figure 3. Cricket score prediction

5. Conclusion

This system is essential for making strategic decisions. It is a holistic approach as it takes in current input from the user. The database maintained is updated on every prediction and the system works efficiently with a massive dataset of two thousands of rows. Our system focuses on only the performance of the player and is very quick at processing due to the clustering of data. Currently the system processes and predicts data for ODI formats as well as on grounds around the world.

References

1. Dost Muhammad Khan, Nawaz Monaurally (July 2010). "Agent Oriented Approach for Implementation of the Range Method of Initial Centroids in K-Means Clustering Data Mining Algorithm" International Journal of Information Processing and Management Volume 1, Number 1,
2. Barr, G. D. I., and B. S. Kantor. (2016). "A criterion for comparing and selecting batsmen in limited overs cricket." Journal of the Operational Research Society 55.12: 1266-1274.
3. Van Staden, Paul Jacobus (2017). "Comparison of cricketers bowling and batting performances using graphical displays."
4. Khabir Uddin Mughal (2015). Top 10 Most Popular Sports In The

- World.<http://sporteology.com/top-10-popular-sports-world/>Accessed 2 February.
5. Machine Learning Group at the University of Waikato (2015). Weka 3: Data Mining Software in Java. <http://www.cs.waikato.ac.nz/ml/weka/>Accessed 12 February .
 6. M. Sipko and W. Knottenbelt (2018). "Machine Learning for the Prediction of Professional Tennis Matches," thesis, Imperial College London.
 7. Laws of cricket (2015). <http://www.lords.org/mcc/lawsofcricket/>Accessed .
 8. Naïve Bayes Classifier (2015). <http://www.ic.unicamp.br/~rocha/teaching/2011s2/mc906/aulas/naive-bayes-classifier.pdf>.
 9. Tejinder Singh, Vishal Singla, Parteek Bhatia (2015). "Score and Winning Prediction in Cricket through Data Mining", International Conference on Soft Computing Techniques and Implementations.
 10. M. Ahmad, A. Doud, L. Wang, H. Hong (2016). "Prediction of rising stars in cricket"
 11. P. Somaskandhan, G. Wijesinghe, L. Bashitha (2017). "Identifying optimal set of attributes that impose high impact on end results of cricket match using machine learning."
 12. M. Rehman, O. Shamim, S. Ismail (2018). "An analysis of Bangladesh One day international cricket: Machine learning approach".
 - A. Tripathi, J. Vankar, B. Vaje, V. Varekar (2006). Cricket Score Prediction system using clustering algorithm.
 13. M. Bailey and S. Clarke (2006). "Predicting the Match Outcome in One Day International Cricket Matches, while the Game is in Progress", Journal of sports science & medicine, vol. 05, no. 04, pp. 480-487.
 14. N. Pathak and H. Wadhwa (2016). "Applications of Modern Classification Techniques to Predict the Outcome of ODI Cricket", Procedia Computer Science, vol. 87, pp. 55-60.
 15. P. Satao, A. Tripathi, J. Vankar, B. Vaje and V. Varekar (2016). "Cricket Score Prediction System (CSPS) Using Clustering Algorithm", International Journal of Current Engineering and Scientific Research, vol. 03, no. 04, pp. 43-46.
 16. <https://web.archive.org/web/20130120040151/>
 17. http://www.iccricket.com/match_zone/historical_ranking.php