

A Review Paper on Cricket Predictions using Various Machine Learning Algorithms and Comparisons among Them

Vraj Chokshi

Student

Faculty of Technology and Engineering, MSU, Vadodara, India

Abstract— With the progression in innovation and in addition in sports, predicting the results of a match has turned out to be so basic. Cricket is one of the most popular and most watched team games in the world. We are predicting the result of a One Day (ODI) cricket matches using machine learning concepts such as supervised learning to predict the champs of the matches. We utilize career statistics and also the team performances such as batting and bowling performances in order to train the models. However, the unpredictable rules governing the game, the capacity of players and different parameters play an essential part in influencing the ultimate result of a cricket match. Therefore we are using supervised learning algorithms to predict the outcome of the game and it will help the coaches of the team to learn and analyse where actually the team is going wrong and the area of the improvement. So in this paper we are using four types of machine learning algorithms and as well as we are comparing with each other in order to get the best results.

Keywords: Supervised Machine Learning, Logistic Regression, Support Vector Machine, Decision Tree, Bayes Point Machine Binary Classification Model

I. INTRODUCTION

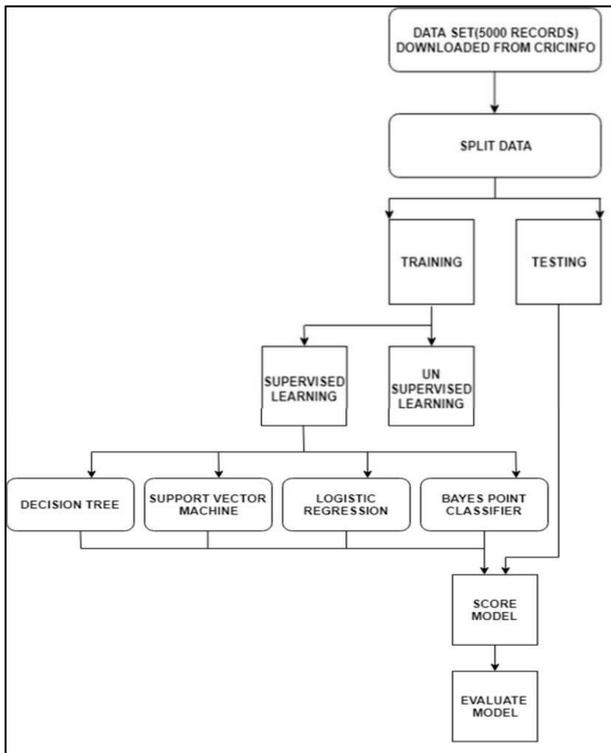
Cricket is one of the most popular sports in this planet. Various natural factors affecting the game, tremendous media scope, and a huge betting market have given strong incentives to model and train the game from various perspectives. However, the complex and different rules governing the game, the ability of players and their performances on a given day, and various other natural parameters play an integral role in affecting the final outcome of a cricket match. As the technology is growing at a faster pace and the huge market in betting and huge demand for cricket has influenced the general population to utilize machine learning calculations to predict the results of cricket matches. Use of machine learning and data science makes life easier in every aspect, using machine learning and predicting the outcomes before the match will allow the players as well as the coaches to analyse the improvement areas. We adopt numerous machine learning concepts and algorithms in order to predict the winner of the cricket matches. And machine learning is booming and machine learning is firmly identified with (and frequently covers with) computational insights, which also focuses on prediction-making through the use of technology. It has solid connections to numerical improvement, which conveys strategies, hypothesis and application areas to the field. Machine learning is some of the time conflated with data mining where the latter subfield concentrates more on exploratory information analysis and is known as supervised learning.

The major contribution of our paper are as follows:

- 1) To predict the winner of ODI cricket matches, we propose a novel dynamic approach to reflect the changes in player Combinations.
- 2) Using machine learning supervised learning algorithms to predict the outcome of the matches.
- 3) Various models in order to check which model has better efficiency and which algorithm gives the accurate or best results.

II. METHODOLOGY

In this section, we explain our approach to the problem in detail, including the definitions and the mechanism of various algorithms used to model the teams in order to predict the cricket matches and to find out the best algorithm. We use four supervised algorithms in order to predict the results of an ODI cricket matches. We have a dataset downloaded from cricinfo website which has almost 5000 records. Each record has its attributes as, the name of the home team, name of the opponent team, match results, toss results, margin by which the team has won, match location, and match date. So, we basically use four and major attributes which contribute to the predicting of the game results, the four attributes are, the name of the home team, name of the opponent team, match results, toss results. Machine learning is firmly identified with (and frequently covers with) computational insights, which also focuses on prediction making through the use of technology. It has solid connections to numerical improvement, which conveys strategies, hypothesis and application areas to the field and machine learning is further divided into supervised learning and unsupervised learning. In this paper we just use supervised learning algorithms such as Logistic regression, support vector machine, Decision tree, and Bayes point machine binary classification model. We use supervised learning algorithm because supervised learning is where you have input factors (x) and an output factor (Y) and you utilize an algorithm to learn the mapping function from the input to the output. The objective is to surmise the mapping capacity so well that when you have new input data (x) that you can predict the output factor(Y) for that data. And we do not use unsupervised learning algorithms in this paper because Unsupervised learning is the place you just have input information (X) and no comparing yield factor The objective for unsupervised learning is to model the underlying structure or distribution in the information so as to take in more about the data. These are called unsupervised learning on the grounds that unlike supervised learning above, there is no right answers and there is no educator or proper teacher. So, this is the reason, we use supervised learning algorithms. We use different supervised learning algorithms such as Logistic regression, support vector machine, Decision tree, Bayes point machine binary classification model.



III. REQUIREMENT ANALYSIS

A. Functional Requirements

1) Product Perspective:

The product is based on Analysing using machine learning techniques such as supervised learning algorithms to predict the outcome of cricket matches.

2) Product features:

Analysing using data mining techniques such as supervised learning algorithms to predict the winners of the ODI cricket matches.

3) User characteristics:

The user is just need to select the two teams to predict the winning team.

4) Assumption & Dependencies:

This mainly depends on the size of the data that is to be analyzed and the type of the model been implemented.

5) User Requirements:

The best data analytics results will be obtained without any limitations, the user is just need to select the two teams to predict the winning team.

B. Non Functional Requirements

1) Efficiency:

Using data mining models are very fast as data mining and machine learning are the new technologies which are used to predict some kind of output with the given data sets as inputs.

2) Reliability:

It is reliable until and unless the data is so huge and complicated and is difficult to process. For the large raw data we have to implement big data to convert that raw data in information.

3) Usability:

It can be used in predictions of ODI cricket matches as they will help the coaches and the players to analyze the area of improvement.

Given the size of the betting industry around the world, there are clearly monetary gains for anyone with access to predominant forecast systems, whether through working with betting companies, selling predictions to professional gamblers or personal betting.

In this report we create machine learning models keeping in mind the end goal to anticipate results of the ODI cricket matches. Principal component analysis was additionally evaluated as an approach to enhance the execution of the models. Use of machine learning and data science makes life easier in every aspect, using machine learning and predicting the outcomes before the match will allow the players as well as the coaches to analyse the improvement areas. For betting industry this model and the probability figure will be very useful for better in deciding which team will win and how much to bet. This will help the team coaches, management to analyze at which situation team needs to improve and can also decide the batting order and bowling changes.

Amount of bet and team can be decided based on this probability and lot of money can be made.

IV. EXPERIMENTS AND RESULTS

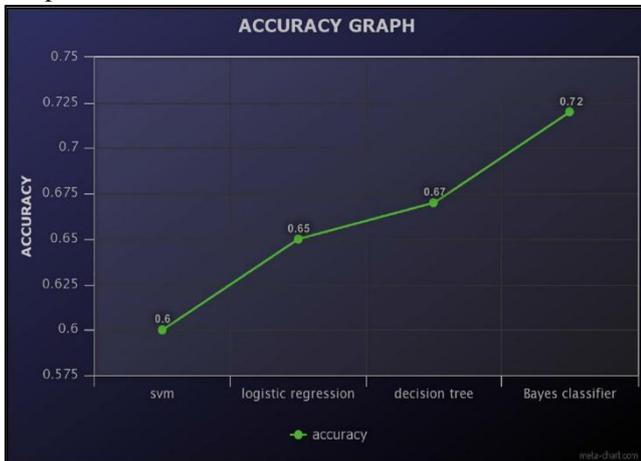
A. Dataset:

To recover all the required statistics, the whole dataset has been scratched from the cricinfo website. The dataset incorporates all the matches played between 19th century and 20th century. The dataset contains the fundamental match points of interest counting the two contending groups, the outcome of the toss, the date when it was held, the venue and the winner of the match for all the matches. Since the effect of the nature on the game can't be anticipated, a sum of 109 matches which were either hindered by rain or ended up in a draw/tie, have been expelled from the dataset. Finally, we divided the dataset into two parts, namely, the test data and the training data. The training dataset contains the 70% of the data from our dataset and the test dataset contains 30% of the data from our dataset. There are a total of 3500 matches in training dataset and 1500 matches in test dataset.

B. Binary Classifiers:

Using various algorithms and numeric features and the results of the matches we assessed an extensive number of binary classifiers utilizing their scikit-learn executions to create supervised classification models, including support vector machine, decision tree, logistic regression, Bayes point machine binary classification. Using the information of future matches to anticipate the result of past matches is foolish. Subsequently, we couldn't complete any kind of cross-approval methodology as it would interfere with the chronological order of the data. The only barrier we faced while evaluating our approach is the inability to compare against previous model due to the different types of datasets used. Our dataset does not have some of the features used by them. For instance, we do not have the details on the timings of the matches (day/night). The results are tabulated in Figure

2. The predominance of our model against the others demonstrates the significance of the combination of various features used. The best algorithm or the model according to our experiment conducted is, Bayes point machine binary classifier as it has gained more accuracy as compared to other models. Other models used in this experiment are also good enough to give accurate results but the accuracy level of the algorithm Bayes point machine binary classifier is more as compared to other models.



V. CONCLUSION

The paper addresses the problem of anticipating the result of an ODI cricket match utilizing the insights of 5000 matches. The curiosity of our approach lies in addressing to the issue as a dynamic one, and utilizing the results of the previous matches as the key element in predicting the winner of the match. We watch that simple highlights can yield exceptionally encouraging results. Predicting the winner of the matches using different supervised algorithms is been achieved and now we can predict the upcoming matches. There might be some more algorithms coming in future which give better results then used in this paper. But the best part is we are now predicting the future events just by using the technology and the computers.

REFERENCES

- [1] Madan Gopal Jhavar, Vikram 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.
- [2] Predicting the Match Outcome in One Day International Cricket Matches, while the Game is in Progress. - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3861745/>
- [3] 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.
- [4] ESPN Cricinfo, <http://www.espn-cricinfo.com>
- [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] Barr, G. D. I., C. G. Holdsworth, and B. S. Kantor. "Evaluating performances at the 2007 cricket world cup." *South African Statistical Journal* 42.2 (2008): 125
- [7] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." *Journal of Machine Learning Research* 12.Oct (2011): 2825-2830.
- [8] Croucher, J. S. "Player ratings in one-day cricket." *Proceedings of the fifth Australian conference on mathematics and computer sport*. Sydney, NSW: Sydney University of Technology, 2000.
- [9] Khan, Mehwish, and Riddhi Shah. "Role of External Factors on Outcome of a One Day International Cricket (ODI) Match and Predictive Analysis."
- [10] Sankaranarayanan, Vignesh Veppur, Junaed Sattar, and Laks VS Lakshmanan. "Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction." SDM. 2014
- [11] Parag Shah (Department of Statistics, H L College of Commerce, Ahmedabad, India) Predicting Outcome of Live Cricket Match Using Duckworth- Lewis Par Score
- [12] Wood, G. H. (1945): "Cricket scores and geometrical progression". *Journal of the Royal Statistical Society Series A*, 108, 12-22.