
REVIEW PAPER ON CRICKET SCORE PREDICTION

Prasad Thorat^{*1}, Vighnesh Buddhivant², Yash Sahane^{*3}

^{*1,2,3}Engineering, Information Technology, VPPCOE & VA, Mumbai, Maharashtra, India

ABSTRACT

With the varied information science approaches in the game of cricket, we can predict the outcomes of cricket matches. The exaggerated monetary advantages have created Cricket a stimulating sport to be subjected to applied mathematics analysis and machine learning. The dynamic nature of Cricket makes the task a difficult one. The varied approaches taken are specializing in ODI matches and therefore the approaches don't seem to be specializing in the dynamic nature of the cricket game. The papers concentrate on the pre-match and post-match predictions. If the demerits of the prevailing work area unit are studied then the loopholes may be exploited. This paper presents an associate analysis of the prevailing work associated with match outcome prediction within the Cricket domain. This paper may be a result of associate in progress analysis, by the tip of the analysis we tend to hope to deal with the drawbacks that may be explored during this paper.

Keywords: Cricket, Machine learning, Linear regression, Naïve Bayes, Winning probability, Random Forest Regression.

I. INTRODUCTION

Cricket is a game played between two teams of 11 players each. The team consists of batsmen, bowlers, and fielders. The batsman plays with the bats and score runs for their team and the bowler bowls the deliveries to the batsman. There are 6 deliveries in each over. There are umpires to keep an eye on the rules and regulations of the game. They make decisions on the field. There are also third umpires who keep an eye on the rules and regulations with the help of technology.

There are 3 formats in i.e. cricket. Test, ODI, and T20. In Test format, the game is played for 5 days and there are 2 innings to each team. The ODI format has 50 overs to each side. The team batting first puts a target of runs in front of the bowling team and the bowling team bats in the second innings to achieve the target put in front of them. The T20 format has 20 overs for each side. The T20 format is the shortest version of the game and has dynamic nature. The game can change very quickly in the T20 format. There are various premier leagues played across the globe. IPL, BBL, CPL are some of the leagues. IPL is the most famous league in the world.

In this paper, there are various sections. Section II deals with the different research papers briefly. In Section III we have analyzed the review papers and also prepared a comparison table to get a clearer view. In Later sections, we have tried to address the demerits of the existing systems and proposed a tentative model which can overcome some of the demerits of the existing works. More focus is on considering the dynamic nature of the game.

II. METHODOLOGY

Research Paper Studied

We reviewed 8 papers from IEEE and we analyzed the reviewed papers. The papers studied were similar to our domain. The papers are based on cricket predictions.

In all the research papers studied it has been found that around 5 papers dealt with the ODI format. Every paper has used some of the common machine learning algorithms i.e. linear regression, random forest regression. In each paper, the algorithms gave different accuracies. Some of the papers also considered multiple models on the same data set just to get better accuracy.

Some of the papers dealt with the squad analysis, some dealt with the winning probability however some dealt with first innings and second innings score prediction, and some papers dealt with sentiments. Using this analysis the motivation is derived.

III. ANALYSIS

The following table shows the comparison between the studied papers.

Table 1: Comparison Table.

Sr. No	Paper Name	Algorithm	Merits	Demerits	Future Scope
1.	Score and Winning Prediction in Cricket through Data Mining Oct 8-10, 2015	1.Linear Regression algorithm 2.Naive Bayes classifier	1.Accuracy of LR is 20% greater than that of the CRR method. 2.Outcome accuracy also increases with an increase in the overs.	1.The prediction is done for the ODI matches. 2.Other factors like toss, ranking of the team, venue, home team advantage are not considered.	1.Accuracy can be increased. 2.Ranking, toss, teams, venue these attributes can be considered while training and testing the model.
2.	Money Ball - Data Mining on Cricket Dataset 2019	1.Naive Bayes classifier 2.Support Vector Machine(SVM) 3.K-Nearest Neighbor method 4.Random Forest method	1.Random Forest algorithm has the most accurate result.	1.Every team has different features/attributes and using those attributes the best scenarios are considered.	1. Accuracy can be increased. 2. Can be used for the tournaments like BBL, CPL, Vitality Blast, etc.
3.	Cricket Match Outcome Prediction Using Tweets and Prediction of the Man of the Match using Social Network Analysis: Case Study Using IPL Data 2018	1.Naive Bayes classifier 2.Support Vector Machine(SVM) 3.K-Nearest Neighbor method 4.Random Forest method 5.Logistic Regression	1.Sentiment analysis has an accuracy of up to 85%. 2.Tweet based and Mixed model has 89% accuracy. 3.Natural parameter-based model has accuracy up to 83%.	1.Sometimes prediction using tweets may go wrong as the game can change very quickly within the span of 2 overs as the span for the tweet-based model is 10 overs.	1.Accuracy can be increased. 2.More tweets and social media inputs can be feed into the models.
4.	Cricket Squad Analysis using multiple Random Forest Regression 2019	1.Linear Regression 2. SVR (Support Vector Regression) 3. Decision Tree 4. Random Forest	1.Provides a mathematical approach to select the players. 2.RMSE value of Multiple Random Forest Regression is greater than LR, SVR, Decision Tree.	1. Considers only ODI matches.	1.Accuracy can be increased. 2.Accuracy can be increased by predicting the matrices of the player against the player.

5.	Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms 2018	1. Linear Regression 2. SVM with linear kernel 3. SVM with polynomial kernel	1.SVM with linear kernel has greater accuracy as compared to LR and SVM with the polynomial kernel.	1.Considers only ODI matches. 2.Accuracy is less.	1.Accuracy can be increased. 2.T20 and TEST format can be included.
6.	Convolutional Neural Network and Feature Encoding for Predicting the Outcome of Cricket Matches 2019	1. Feature Encoding 2. Shallow CNN architecture 3. Random Forest 4.Linear SVM Classifier	1.Predicts outcome even before the match starts.	1.Due to fewer entries and attributes deep CNN can not be used. 2.Accuracy is less.	1.Accuracy can be increased. 2.T20 and TEST format can be included.
7.	Data Mining and Machine Learning in Cricket Match Outcome Prediction: Missing Links 2019	i)Naive Bayes ii)Linear Regression 2. Simulation-based approach 3.Team composition approach 4.Collective knowledge approach	-----	1.Dealing with lack of data about certain players. 2.Not giving enough consideration to the conditions the game is played in 3.Taking an ensemble approach	1.Cricket match outcome prediction remains a new and promising research area. Due to the complex and dynamic nature of the game, achieving high accuracy scores remains a challenging task, especially in the T20 format.
8.	Outcome Prediction of ODI Cricket Matches using Decision Trees and MLP Networks 2018	Multilayer Perceptron Network Decision Trees	1.MLP Networks are capable to learn & train in real-time using partial fitting property 2.Decision trees can handle both numeric as well as categorical data and also multi-output problems.	1.MLP Networks are highly sensitive for feature scaling. 2.Decision Trees are usually very unstable and even small modifications in the data might lead to an entirely different tree being generated.	1.This analytical study can be expanded further in terms of the team composition perspective. This methodology and technique can also be applied to predict the outcomes of games like hockey and football.

IV. RESULTS AND DISCUSSION

a) Motivation

We found that in almost every study it is mentioned that the prediction is done for the ODI cricket matches but we are going to develop a system that will predict the score of the first innings of T20 matches. T20 format is far different from the other formats, here the game changes within the span of an over, so we aim to make a

dynamic system that will be having the prediction outcome probability higher even if this format is changing ball to ball.

We found that the Linear Regression model and Random Forest model gave the greater probability of the prediction so we are going to use these models for prediction.

b) Tentative system architecture

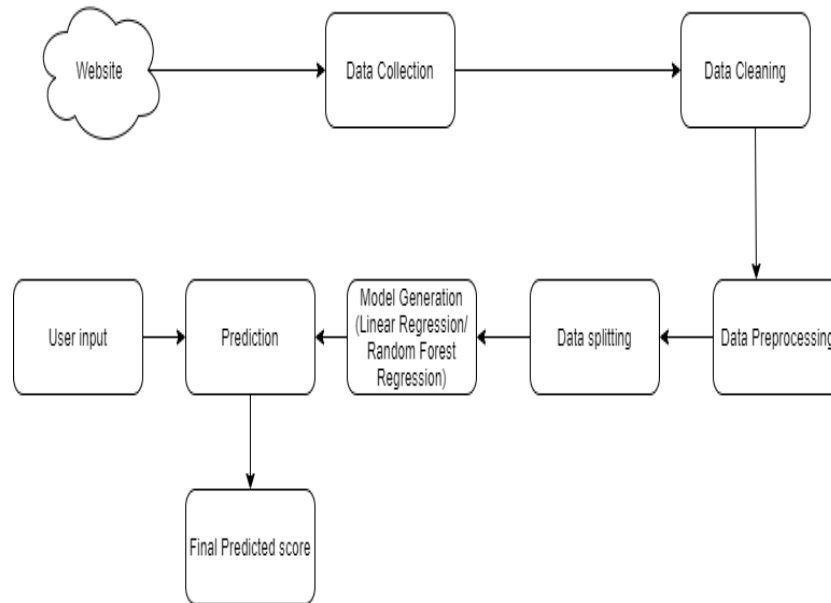


Figure:1 Cricket Score Prediction Architecture.

V. CONCLUSION

After studying the research papers, we found that most of the papers have only considered the ODI format. There is a need to focus on the T20 format. The T20 format is a dynamic format and the parameters for prediction may vary. Linear Regression and Random forest algorithms are preferred by most of the researchers. These algorithms gave the highest accuracies. The researchers have not considered the dynamic feature of the T20s.

Now after studying these papers, we are planning to create a model which can make a better prediction for the T20 (IPL) game with higher accuracy.

VI. REFERENCES

- [1] Tejinder Singh, Vishal Singla, Parteek Bhatia; - Score and Winning Prediction in Cricket through Data Mining; Oct 8-10, 2015
- [2] D. Thenmozhi, P. Mirunalini, S. M. Jaisakthi, Srivatsan Vasudevan , Veeramani Kannan V, SagubarSadiq S; Moneyball - Data Mining on Cricket Dataset; 2019
- [3] A.N.Wickramasinghe, Roshan D.Yapa; Cricket Match Outcome Prediction Using Tweets and Prediction of the Man of the Match using Social Network Analysis: Case Study Using IPL Data; 2018
- [4] Nigel Rodrigues¹, Nelson Sequeira², Stephen Rodrigues³, Varsha Shrivastava⁴; Cricket Squad Analysis using multiple Random Forest Regression;2019
- [5] Animal Islam Anik, Sakif yeaser, A.G.M. Emam Hussain, Amitabha Chakraborty; Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms;2018
- [6] Siyamalan Manivannan, Mogan Kausik; Convolutional Neural Network and Feature Encoding for Predicting the Outcome of Cricket Matches;2019
- [7] Manuka Madranga Hatharasinghe, Guhanathan Poravi Data Mining and Machine Learning in Cricket Match Outcome Prediction: Missing Link;2019
- [8] Jalaz Kumar, Rajeev Kumar, Pushpender Kumar; Outcome Prediction of ODI Cricket Matches using Decision Trees and MLP Networks;2018