

# Extrapolation of One Day International (ODI) Cricket Match's Consequences

Kirupa Shankar K M, Alok Kumar

Assistant Professor, Department of Computer Science and Engineering, Government College of Engineering,  
Dharmapuri, Tamil Nadu, India

## ABSTRACT

A great many fans devour data identified with cricket each day, through media or by watching games. Such expectation action includes a substantial thinking mental interaction that generally summarizes an entire collection of disseminated data, (for example, measurements on players, individual player's exhibition) to figure out who will win and who will lose. In this paper an exemplary has been projected to predict the result of the match. for example expectation of One-Day International (ODI) cricket match result for Indian group in contradiction of all international resistances has been introduced. A great deal of effort has gone into gathering raw data and filtering it for a variety of aspects that might influence the outcome of an ODI cricket match. A few extraordinary methodologies embraced on behalf of database arrangement and characterization model discovering that permit one to foresee the match result with 92% exactness which is far more noteworthy than the slog recently shown. Different Data Mining calculations were applied on various sizes of preparing and testing informational indexes. It was discovered that k-Nearest Neighbors (KNN) outperformed three other well-known characterization calculations (for instance Decision Tree, Naive Bayes, and Linear Regression). The expectation model can be utilized to profit Board of Council for Cricket in India (BCCI) by surveying the benefits of specific systems of play. Besides, cricket examiners, media can likewise utilize the model for pre-match investigation.

**Keywords :** Data Mining, Decision Tree, K Nearest Neighbors (KNN), Linear Regression, Naive Bayes, Winning Probability

## I. INTRODUCTION

In cricket, the batter attempts to score runs by hitting the bowled balls as they roll towards him. The roller (bowler) then tries to get the batter out of the game once more. There are several criteria that must be followed in order for the batsman to get caught by the rollers or the defenders. Every hitter continues to bat until he is out of the game. Similarly, the flapping group's turns are finished when either the 10 batsmen are out or the handling group has bowled 50 overs; nevertheless, the batting group now has the opportunity to bowl while the bowling crew is shot at batting. The winning position of the team can be

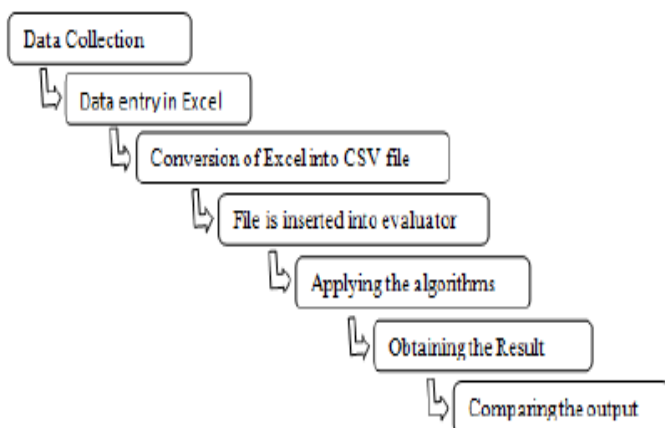
predicted by the scores by which a team secured most runs in the match.

In this paper a prototypical has been anticipated which predicts the result of the match. For example, against all international regulations, the expectation of a One-Day International (ODI) cricket match outcome for the Indian squad has been created. A great deal of effort has gone into gathering raw data and filtering it for a variety of aspects that might influence the outcome of an ODI cricket match. A few extraordinary methodologies embraced for database development and order model discovering that permit one to anticipate the match result with 85% precision that is to say a far more prominent than

the work recently shown. On varying scales of developing and evaluating informational collections, several Data Mining calculations were used. It has been discovered that k-Nearest Neighbors (KNN) has beaten three more famous calculating methods. (For example Choice Tree, Naive Bayes, Linear Regression). The forecast model can be utilized to profit Board of Council for Cricket in India (BCCI) by surveying the benefits of specific systems of play. Besides, cricket examiners, media can likewise utilize the model for pre-match analysis. The matches of ODI which was played between the nations from 2015 to 2019 were used to compile the historical data. The important point to note here is the all the matches played during the aforementioned duration have been taken into consideration.

## II. RELATED WORK

Few people have functioned on accurately envisaging the outcome of an ODI match. "Winning Predicting" is one such slog. The model that we are making, as far as the yield created that is, foreseeing the last winning probability in the match. In this method, information mining techniques such as Linear Regression, Nave Bayes Classifiers, Decision Tree, and K Nearest Neighbors were applied to a database of ODI matches played from 2011 to 2019.



**Fig.1 : Work Methodology**

## III. METHODOLOGIES AND TOOLS USED

Various data mining techniques are utilized in our education to envisage the consequences of a cricket match. The R programming language is used to perform classification algorithms and make estimations. The outcomes of the techniques are examined with the following four classification algorithms.

### Classification algorithm

In ML (Machine Learning) algorithm, arrangement refers to a technique for determining which session another example belongs to, based on a preparatory set of data that includes impressions with established class membership. For example, based on its characteristics or highlights, relegating an obscure specimen of a bloom to the specified flower species "Lotus" or "Hibiscus" in view of its attributes or highlights. Grouping is an administered learning wherein a preparation set of accurately recognized and the number of occurrences is provided. A classifier is a computation that is used in the execution of grouping. Linear Regression, Nave Bayes classifier, Decision Tree, and K Nearest Neighbors were used as separators in this study.

### Linear Regression

The information supplied here for the class and the properties are said to be numeric. And that too this is on the off chance. The information is direct relapse classifier and it is utilized for order. The information about a class and its characteristics is numerical, and the order is determined by a direct relapse classifier. This is a common measuring technique. Given a dataset, relapse is used to predict a range of numeric characteristics (also known as uninterrupted qualities).

$$X = l^0 + l^1 \cdot a^1 + l^2 \cdot a^2 + \dots + l^v \cdot a^v$$

Where,

$x \rightarrow e$  class;

$a^1, a^2, \dots, a^v \rightarrow$  values of the attributes;

$l^0, l^1, \dots, l^v \rightarrow$  loads.

The loads can be figured from the preparation information. Here, the documentation gets somewhat weighty, in light of the fact that a method of communicating the trait esteems is required for each preparation case.

### Naïve Bayes classifier algorithm

Separator based on probabilities The Bayes' hypothesis, which is based on considerable (naive) freedom suspicions, is the basis for the Naive Bayes classification approach. This separator accepts that occurrence (or nonappearance) of a specific component of a class is autonomous to the occurrence (or nonattendance) of some other element. It deliberates these structures to autonomously add in order to determine the likelihood, regardless of whether the qualities are dependent on one another. Be contingent on the exactness of the possibility, the Naive Bayes separator is used to efficiently construct a controlled learning environment.

Utilizing Bayes' theorem, the restrictive probability is given by,

$$P(A/B) = (P(B/A) P(A))/P(B)$$

Where,

$P(A/B) \rightarrow$  given prognosticator, the class's posterior possibility (attribute),

$P(A) \rightarrow$  previous class probability,

$P(B/A) \rightarrow$  The probability of a prognosticator for a particular class is called likelihood.

$P(B) \rightarrow$  prognosticator's previous likelihood Decision Tree

One of the most often used inductive derivation algorithms is the decision tree. It is a capacity shut to discrete values and furthermore can be blessed to receive a boolean capacity. It is an inductive learning calculation basic on the case, which can be ordinarily used to shape an arrangement and expectation models. In different manners to address arrangement problems, decision tree are usually utilized as a method. It is a prescient displaying strategy for characterization, grouping and prediction, which partition the pursuit space into various subsets through "divide and rule".

A tree is needed to work in support of the demonstrating for the grouping interaction utilizing this method. As soon as the tree is built, a finite ordered list in informational collections is pragmatic and gets the classification outcome. In the decision tree, there are two fundamental steps: build a tree and put the tree into the informational index. . It utilizes Entropy along with Fact benefits to develop a decision tree.

### Entropy

1. Entropy for single attribute

$$E(s) = \sum_{i=1}^c -p \log_2 p_i$$

2. Entropy for double attribute

$$E(T,X) = \sum_{c \in X} P(c) E(c)$$

### Fact benefits

$$\text{Benefits}(T, X) = \text{Entropy}(T) - \text{Entropy}(T, X)$$

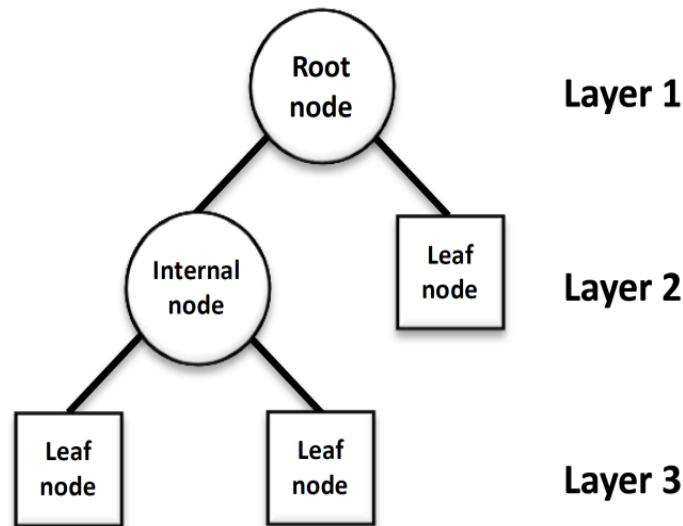


Fig. 2 : The structure of Decision Tree

Table. 1 : Precision and Recall Comparison

Metrics	Linear Regression		Naïve Bayes		Decision Tree		KNN	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
India	0.706	0.800	0.929	0.867	1.000	0.933	1.000	0.922
New Zealand	0.500	0.375	0.778	0.875	0.973	1.000	0.956	1.000
Weighted Avg.	0.634	0.652	0.876	0.870	0.921	0.920	0.923	0.921

**B. K Nearest Neighbors(KNN)**

The KNN algorithm is a classifier algorithm. This algorithm practices a distance routine to forecast a new class. The computation is consistently based on the Euclidean remoteness in the middle of a test and the prescribed preparation tests. To choose which of the focuses from the preparation dataset are sufficiently comparable to be viewed as while picking the session for ground breaking perception is to choose the ‘k’ nearest information focuses, and to yield the utmost well-known class among these. As a result, a model is defined by the majority vote of its

neighbours. K-nearest-neighbour is an adaptable calculation, utilized in an immense sum of fields. Complacent recovery, quality articulation, protein-protein cooperation and 3D construction expectation lie in scarcely any extraordinary and non-minor utilizations of kNN.

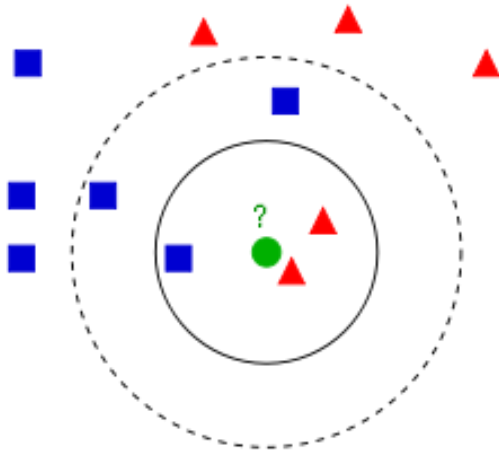


Fig. 3 : Example for kNN

**Distance Function**

The Euclidean function is given,

$$d(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2}$$

**INFORMATION GATHERING AND PROVISIONING**

The information has gathered from <http://www.espnricinfo.com>. In this website information of the relative multitude of matches are accessible freely. The dataset comprises of absolute matches barring all the downpour hindered and downpour stranded games, played somewhere in the range of 2011 and 2019.

Table. 2 : Descriptions of ODI attributes

ATTRIBUTES	DESCRIPTION
Batting Team	The team which is as of now batting.
Present Score	The present score of the game for the batting team subsequently specific overs either in first innings or in quite a while.

Overs	The quantity of overs rolled at the specific phase of batting team.
Groove	The last groove of the team toward the finish of primary innings.
Target	The objective given to the team batting in the subsequent innings.

**INVESTIGATION AND EXAMINATION OF PARTICULARS**

The specifics (datasets) were studied in R, which includes several machine learning calculations for implementing the information withdrawal measure. The particular has been allocated independently into preparing with a group of tests that belong to specific tasks or feature as the ODI counterparts played till 2018 2019 separately.

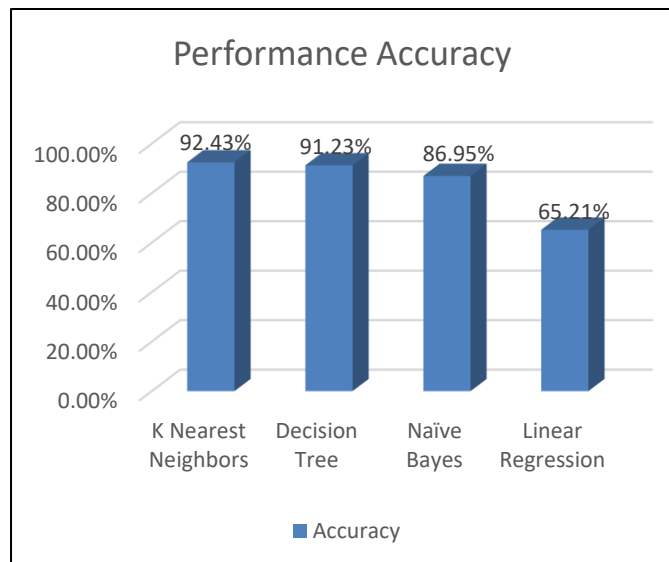
**IV. RESULTS AND ANALYSIS**

This part shows the aftereffects of different characterizations molded by R and examine the outcome with ongoing works in anticipating the result of the match. The outcomes have been gotten and noticed the groupings, exactness and grids. The exhibition of those characterizations have analysed and displayed in Table 1. Here we have taken two groups to anticipate the result of the match. One is India and another group is New-Zeland. There are several Metrics that are used to analyze a match's outcome. Table 2 has four computations that may be compared.

**Table. 3 : PERFORMANCE COMPARISON**

Classifiers	Linear Regression	Naïve Bayes	Decision Tree	KNN
Root Mean Square Error	0.589	0.239	0.180	0.140
Correctly Classified Instances	65.217%	86.95%	91.23%	92.433%
Incorrectly Classified Instances	34.783%	13.04%	8.77%	7.567%

By looking table-3 KNN has most noteworthy forecast exactness by 92% followed by Decision Tree by 91%. Next Naïve Bayes by 87%. Finally the strategy which is lower expectation precision is Linear Regression by 65%. Correlation of execution precision is displayed in underneath.



**Fig. 4 : Comparison of Performance Accuracy**

### V. CONCLUSION

Four supervised data mining methods were employed to the data set in this study to predict the outcome of

a cricket match, and their investigative accuracy was assessed. The findings show that the KNN algorithm performs the best, with a 92 percent accuracy, indicating that it is a potentially useful and efficient classification technique. A comparison of all four Classifiers using the R programming language is also performed, with KNN proving to be the best in this case as well. As a result, the performance of KNN is superior than that of other classifiers. A graph of model performance is also drawn. The BCCI and the media may benefit from this research, which will serve as a foundation for deciding on specific assistance for them. However, this approach may be utilised to assist the Board of Control for Cricket in India (BCCI) in predicting the match winner.

### VI. REFERENCES

- [1]. Abdelhalim, Nakkabi Y, Amany, Traore I. Creating Decision Trees from Directions using Rule Based Decision Tree-I. Computational Intelligence 2016; 32: 216-239.
- [2]. Ramasubramanian K, Singh A. Machine Learning Using R. Apress 2017.
- [3]. Naveed Nadeem, Ahmed Akhter, Talha Ahmed Mirza Osama Tariq Mahmood. Envisaging the cricket performance concluded data mining. IEEE International Conference On Computer and Emerging Technologies 2013.

- [4]. Baobao W, Minru S, Jinsheng M. Information Gain and Extension Relativity are used to improve the K Nearest Neighbor Algorithm. International Conference on Condition Monitoring and Diagnosis 2008; 1314-131
- [5]. Cai Z , Jiang L, Jiang S, Wang D. Improving KNearest-Neighbor for Classification: A Survey. Fourth International Conference on Fuzzy Systems and Knowledge Discovery 2007; 679-683.
- [6]. Gurpreet Singh Josan, Kaur, Manpreet Singh, Parneet. Data mining techniques based on classification and prediction to forecast slow learners in the education industry. Proceedings in ComputerScience 2015; 500-508.
- [7]. Edin Osman, Mirza. Predicting Student Performance Using Data Mining. Journal of Economics and Business 2012; 1-10.
- [8]. Gill P S, Muthukumaran S, Swartz T B. One-day cricket modelling and simulation. Canadian Journal of Statistics 2009;143-160.
- [9]. Padma P, Raj K . An example of association rule mining in action: Team India. International Conference on Computer Communication and Informatics 2013; 1-6.
- [10]. Gawali, Rajendra D, Vaibhav P, Vasani. Data mining techniques are used for classification and performance evaluation . International Journal of Innovative Research in Science, Engineering and Technology 2014.
- [11]. Lantz B. Machine learning with R. Packt Publishing Ltd 2015.
- [12]. Gupta C, Gureja D, Goyal K, Taneja S. A Review of K Nearest-Neighbor Techniques for Data Classification. International Conference on Computing, Informatic and Network 2014: 69-113.
- [13]. <https://espncriinfo.com>. Cricket Data's
- [14]. Junaed Sattar, Laks V S Lakshmanan , Vignesh Veppur Sankaranarayanan. Autoplay: An method to ODI cricket simulation and prediction based on data mining. Society for

Industrial and Applied Mathematics (SIAM) International Conference on Data Mining 2014:1064–1072.

#### Authors :



Mr. KIRUPA SHANKAR K M is working as an Assistant Professor in the Department of Computer Science and Engineering at Government College of Engineering, Dharmapuri, Tamilnadu. His research interest includes Knowledge in Data Mining, Machine Learning, and Cloud Computing.

Address: Government College of Engineering, Dharmapuri, Tamilnadu - 636704



Mr. ALOK KUMAR is working as an Assistant Professor in the Department of Computer Science and Engineering at Government College of Engineering, Dharmapuri, Tamilnadu. His research interest includes Knowledge in Data Mining, Machine Learning, and Deep Learning.

Address: Government College of Engineering, Dharmapuri, Tamilnadu – 636704

#### Cite this article as :

Kirupa Shankar K M, Alok Kumar, "Extrapolation of One Day International (ODI) Cricket Match's Consequences", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 3 Issue 3, pp. 2157-2163, March-April 2018. Available at doi : <https://doi.org/10.32628/CSEIT182015>  
Journal URL : <https://ijsrcseit.com/CSEIT182015>