ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2020.0313

PREDICTION ON IPL DATA USING MACHINE LEARNING TECHNIQUES IN R PACKAGE

G. Sudhamathy and G. Raja Meenakshi
Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, India

Abstract

One of the most exciting outdoor games that reached everyone heart is cricket. There are several series held and one among that created a magnificent history in the arena of sports is Indian Premier League (IPL). It has reached its popularity with successful brand in the world of sports and usually will be conducted among 8 teams. This proposed paper is specifically concentrating on enactment and measuring the difference between the models to foretell the captivating team of an IPL match. Data is accessed by the computer programs developed using Machine learning to build models. As of now, data analysis is need for each and every fields to examine the sets of data to extract the useful information from it and to draw conclusion and as well make decisions according to the information. The algorithm first analyses the data to create a model, specifically for understanding the patterns or trends. For creating the mining model, the model is optimized by selecting parameters and iterating. To extract actionable patterns and detailed statistics, the parameters are then fed into the dataset. This work focuses on finding the meaningful information about the IPL Teams by using the functions of R Package. R reduces the complexity of data analysis as it displays the analysis results in the form of visual representations. The dataset is loaded and a set of pre-processing is done followed by feature selection. Four machine learning algorithms Decision Tree, Naive Bayes, K-Nearest Neighbour and Random Forest are applied and the results are compared to measure the accuracy, precision, recall and sensitivity. The best of the four machine learning techniques is then applied to predict the winner and visualizes the results as graphs.

Keywords:

Prediction, IPL, Machine Learning, R Package

1. INTRODUCTION

England first introduced T20 Cricket in 2003. Because of its shorter format, it became very popular. Due to its popularity of high voltage action, T20 came to India also. BCCI initiated a 20-20 cricket tournament Indian Premier League (IPL) in 2008. BCCI has been organizing the IPL T20 cricket tournament every year. The use of analytical methods in various aspects of cricket including results prediction is very important. There is a huge demand for the algorithm that best predicts the result of cricket because of its popularity and huge amount of money involved in the game. Thus the analysis of IPL results becomes more important. Prediction of outcome of a match using machine learning algorithms is an important aspect in cricket. Records of the past performance of players and other related data can be analysed to create models that predicts the winning team. This model can be created using the machine learning algorithms such as Decision Tree, Naive Bayes and K-Nearest neighbour and their results can be compared based on the Evaluation Measures as accuracy, precision, recall, sensitivity and error rate.

The proposed paper organized as follows. In section 2, various works in the field are discussed and the gap in exploring using machine learning techniques available in R has been highlighted. Section 3 discusses the methodology of the approaches applied in this paper using a block diagram. Results and discussions are detailed in the section 4. This section explores the results for better understanding. It is also important that the performance metrics derived from the models is proving the high accuracy and efficiency of the built model. Section 5 concludes the work done in this paper.

2. LITERATURE REVIEW

The work done on Data Mining of Cricket dataset describes the various data mining techniques viz Decision Tree, Naive Bayes, KNN, Random Forest applied on the IPL dataset, the model is built for predicting the results of the matches. The best attributes were selected using the Wrapper and Ranker method and then the classification has been done. This work was done with the help of WEKA.

Gupta et al. [2] says that the selection of the best team is always required by the management for best outcome. The paper provides the optimal solution to select the best team using Data Mining Techniques rather than following the traditional method which is tedious. When we are declaring a time for the particular championship it is mandatory to select the best team and so the chance of the team to be the champion becomes easy.

In [3], the authors proposes the fuzzy clustering logic. The results of the IPL batting Statistics were grouped into various clusters and it gave efficient and effective accurate results with the Data Mining Technique – Clustering. This work has been done with the help of MATLAB. The concept of clustering is used in order to classify batting statistics of the Indian Premier League which has the fuzzy data into appropriate clusters.

Raza Ul Mustafa et al presented a study [4] on the investigation of the feasibility of using the Twitter data to forecast the results of the match. The work has been proposed to check the machine learning techniques’ effectiveness when applied on data collected to derive insight obtained from social media networks and other real world events are predicted. The techniques used in their work are Support Vector Machine, Naive Bayes Classifier and the Linear Regression. The SVM technique holds good.

Live Cricket Score and Winning Prediction work [5] describes about the building of the model which predicts the score for the chasing team and will estimate the score of the second innings of match. The proposed work uses the concepts of Linear Regression, Naive Bayes Classifier and Reinforce Learning Algorithm. The factors such as toss result, ranking of the team, home team advantage were considered.

Sankaranarayanan [6] gives the idea about building a system of prediction that takes the historical data and predicts the victory or loss of the forthcoming matches.

They used Linear Regression, Nearest Neighbouring and clustering methods will present the mathematical results will exhibit the performance of the algorithm for predicting the results of the model.

Parag Shah, in his work of predicting outcome of the live match [7] proposed model which predict the match result after each ball. The par score concept has been used Duckworth & Lewis, the probability is calculated and it provides clarity of who will win the match

Kaluarachchi [8] used artificial intelligent technique specifically bayes classifiers in machine learning to classify the factors as home game advantage, day/night effect the toss and batting first that affects the result of the match. The final outcome of this work is the delivered as software tool called CricAI. The tool gives the probability of winning based on the input factors such as home game advantage is available at the beginning of the match. The CricA1 can be used in real-world applications when teams are playing cricket. It is used for modifying certain factors to increase the chances of winning in the real field.

The models use Data Analytics methods from machine learning domain. In a rain affected match the prediction of the result may be difficult .In a rain affected match Batting, Bowling, Fielding, Team Selection, Result Prediction, Target Revision is very important. The match prediction can be solved using the mathematical model created based on the insights of the results of prior matches. The Predictor models are created with the help of SVM. This work is proposed using Deep Mayo Predictor [9].

The software developed based on the work by Jayshree Hajgude [10] for Statistical Analysis and Data Mining forms a dream team with the Bayesian Prediction Technique and Parameter based filtering. The database has the details of the current IPL players. This work helps to mine the needed data for using by prediction algorithm in order to obtain the statistical analysis of each player.

Predictive model is developed to predict the cricket score and player performance using ODI dataset [11] Performed Supervised methods like SVM, Naïve Bayes. Clustering methods like KNN and MLP method to classify accurately.

The work done by Kaluarachchi [8] used artificial intelligent technique specifically bayes classifiers in machine learning to classify the factors as home game advantage, day/night effect the toss and batting first that affects the result of the match. The final outcome of this work is the delivered as software tool called CricAI. The tool gives the probability of winning based on the input factors such as home game advantage is available at the beginning of the match. The CricA1 can be used in real-world applications when teams are playing cricket. It is used for modifying certain factors to increase the chances of winning in the real field.

The models use data analytics methods from machine learning domain. In a rain affected match the prediction of the result may be difficult. In a rain affected match batting, bowling, fielding, team selection, result prediction, target revision is very important. The match prediction can be solved using the mathematical model created based on the insights of the results of prior matches. The Predictor models are created with the help of SVM. This work has been proposed using Deep Mayo Predictor [9].

The software developed based on the work by Hajgude [10] for Statistical Analysis and Data Mining forms a dream team with the Bayesian Prediction Technique and Parameter based filtering. The database has the details of the current IPL players. This work helps to mine the needed data for using by prediction algorithm in order to obtain the statistical analysis of each player.

Predictive model is developed to predict the cricket score and player performance using ODI dataset [11] Performed Supervised methods like SVM, Naïve Bayes. Clustering methods like KNN and MLP method to classify accurately. IPL match winner prediction is carried out using ML techniques [1]. IPL is interesting game each year there is lot of expectations of who will win the prestigious title. IPL is game where the result can be changed in just few seconds so to predict the winner ML algorithms like SVM, Logistic Regression, Naïve Bayes, Decision Tree and KNN.

To overcome these existing modules, we developed a predictive technique for predicting the winning team.

3. METHODOLOGY

The proposed method consists of five sub modules, namely, loading the dataset, pre-processing, feature selection, classification using various algorithms and comparison of algorithms as shown in Fig.1.

3.1 LOADING THE DATASET

The dataset name is matches.csv (IPL Matches data from 2008 to 2017) whose size is 117,096 bytes and it is taken from the Kaggle Repository. The number of attributes is 18 and total number of records is 637. The Attributes of the dataset is id, season, city, date, team1, team2, toss_winner, toss decision, result, dl_applied, winner, win_by_runs, win_by_wickets, player_of_match venue, umpire1, umpire2, umpire3. The dataset is loaded into the R Tool and command read.csv() is used to upload the data and this data is stored in the dataset named IPL data.

3.2 DATA PRE-PROCESSING

Data Pre-Processing plays a vital role in machine learning. It transforms raw data into a useful data format. Commonly it is used as a preliminary step to clean the data. Data Pre-Processing transforms the data into a format for more easily and error free processing for the classification. The dataset is first processed to remove the null attributes and the records that contain the NA attributes. The attribute umpire3 is removed initially as it had no values. The fields date and player_of_match are converted to numeric fields. Records with NA in the winner and player_of_match are removed. The levels in the winner fields are also dropped to make it a non-factor variable. These pre-processing has to be done before the feature selection and classification techniques.

3.3 FEATURE SELECTION

Feature selection is the use of specific attributes in the dataset to maximize efficiency Feature selection is also known as variable selection. It is important phase in machine learning because it significantly improves the performance by eliminating redundant and irrelevant features and also at the same time speeding up the learning task. Feature selection is done using two functions namely the Boruta() and the importance() functions. The Boruta() function is in the Boruta package and the importance() function is in the randomForest package. The Boruta function a narrow – down search for relevant features by comparing with original attributes. The importance is achievable at random estimated using their permuted copies, and progressively eliminating all irrelevant features stabilize that test. The importance() function is the function of extraction for variable importance measured as produced by random Forest. With the Boruta() function, date, dl_applied, umpire2 are confirmed as unimportant. With the importance() function, umpire1, umpire2, venue, result and dl_applied are with least Mean Decrease Accuracy. Hence, the fields umpire1, umpire2, venue, dl_applied and result were removed by comparing both the algorithms.

3.4 CLASSIFICATION

In Machine Learning, classification is an important technique to classify different classes. It is a supervised learning method in which the computer program learns from the training data, and uses this learning to classify new data. Here four different classification algorithms are applied, namely, Decision Tree, Random Forest, Naive Bayes and K-Nearest Neighbour.

3.4.1 Decision Tree:

Decision Tree is one of the supervised learning algorithms which is used for both classification and Regression. A Decision Tree is a graph that uses a Tree based method to illustrate every possible outcome of a decision. A decision tree is a decision support tool which uses a tree-like structure and their possible consequences. The packages caret and rpart.plot are used from which the functions rpart(), createDataPartition(), trainControl(), train(), prp() and predict() are function are used to get the result of decision tree algorithm.

3.4.2 Random Forest:

Random forest is a supervised learning method. In the random forest classifier, the more the number of the trees the more the best accuracy for the model. Random Forest is also an ensembled based method used for classification, regression and other tasks. The package randomForest is used which contains the functions sample(), randomForest() and plot() that are used to obtain the results of the Random Forest algorithm.

3.4.3 Naive Bayes:

A Naive Bayes classifier is a supervised learning algorithm which works based on Bayes' theorem. Naive Bayes classifiers uses conditional probability theorem to classify the data. Bayes classifiers will assume whether strong or naïve independence between attributes of data points. These classifiers are broadly used in text categorization based problems because they are easy to carry out. Naive Bayes is also known as independence Bayes. The naïve Bayes () function in the naïve Bayes Package of R is used to obtain the results.

3.4.4 K-Nearest Neighbor:

KNN is the simplest classification algorithm. A k-nearest-neighbour is a classification algorithm that attempts to determine how near the group of data points are around it. Used for both classification and regression base problems. The package RWeka and the function IBk() are used to achieve the results of this algorithm.

3.5 COMPARISON OF CLASSIFICATION ALGORITHMS

The selection of the best classification algorithm for a given dataset is important to acquire the best result. It is a complex one, because it requires to make several important methodological choices. In this work the focus is on the measures used to assess the classification performance and rank of the algorithms. The top most popular measures are presented here and their properties are discussed. Numerous measures have been proposed over the years.

In the field of machine learning, a confusion matrix, is known as error matrix that often visualizes the performance of a classification models .Each row of the matrix will represent the samples of the predicted class. In predictive analytics, confusion matrix is a table with 2*2 Matrix that reports the number of false positives, false negatives, true positives and true negative in total samples. Performance measures are accuracy, true negatives precision, recall, sensitivity, specificity and error rate.

All the above said performance measures are based on the Positive Class (P), Negative Class (N), True Positive (TP) samples, True Negative (TN) samples, False Positive (FP) samples and False Negative (FN) samples.

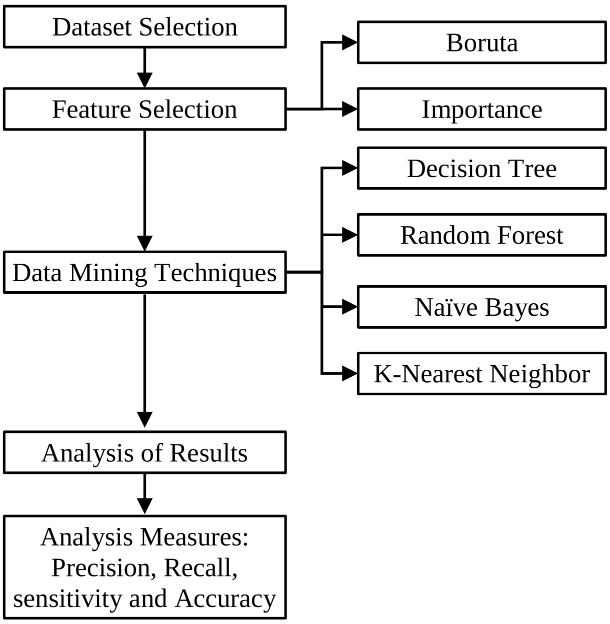


Fig.1. Methodology Diagram

P: Positive Class = True Positive + False Negative = Samples predicted as CSK

N: Negative Class = FP + TN = Samples predicted as Non-CSK

Table.1. Prediction Results

Actual / Predicted		**Predicted**	
		CSK	**Non-CSK**
Actual	**CSK**	TP	FN
	Non-CSK	FP	TN

Based on the above factors performance measures of the classifiers are discussed

Accuracy is a measure that calculates the rate of correct classifications.

$$\text{Accuracy} = TP + TN / P + N$$

Precision is ratio of positives among the total number of instances.

$$\text{Precision} = TP / TP + FP$$

Recall is the ratio of true positives among the true positives and false negative instances.

$$\text{Recall} = TP / TP + FN$$

Sensitivity is a ratio of positive classes that are correctly identified as positive.

$$\text{Sensitivity} = TP / P$$

Specificity is a ratio of negative classes that are correctly identified as negatives.

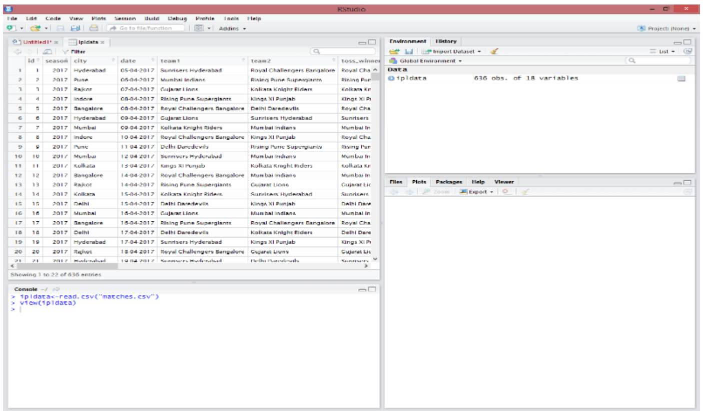
$$\text{Specificity} = TN / N$$

Error rate is rate that measures the inaccuracy predictions the classification algorithm.

$$\text{Error Rate} = FP + FN / P + N$$

4. RESULTS AND DISCUSSIONS

The dataset is loaded into the R Software using the function read.csv() and pre-processing is done. Boruta performs a narrow-down search. It works based on extracting best features by comparing original attribute's importance. It eliminates all irrelevant features to stabilise the performance of the test.



id	season	city	date	team1	team2	toss_winner
1	2017	Hyderabad	05-04-2017	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Cha
2	2017	Pune	06-04-2017	Mumbai Indians	Rising Pune Supergiants	Rising Pur
3	2017	Rajkot	07-04-2017	Gujarat Lions	Kolkata Knight Riders	Kolkata Kr
4	2017	Indore	08-04-2017	Rising Pune Supergiants	Kings XI Punjab	Kings XI P
5	2017	Bangalore	08-04-2017	Royal Challengers Bangalore	Delhi Daredevils	Royal Cha
6	2017	Hyderabad	09-04-2017	Gujarat Lions	Sunrisers Hyderabad	Sunrisers
7	2017	Mumbai	09-04-2017	Kolkata Knight Riders	Mumbai Indians	Mumbai In
8	2017	Indore	10-04-2017	Royal Challengers Bangalore	Kings XI Punjab	Royal Cha
9	2017	Pune	11-04-2017	Delhi Daredevils	Rising Pune Supergiants	Rising Pur
10	2017	Mumbai	12-04-2017	Sunrisers Hyderabad	Mumbai Indians	Mumbai In
11	2017	Kolkata	13-04-2017	Kings XI Punjab	Kolkata Knight Riders	Kolkata Kr
12	2017	Bangalore	14-04-2017	Royal Challengers Bangalore	Mumbai Indians	Mumbai In
13	2017	Rajkot	14-04-2017	Rising Pune Supergiants	Gujarat Lions	Gujarat Lic
14	2017	Kolkata	15-04-2017	Kolkata Knight Riders	Sunrisers Hyderabad	Sunrisers
15	2017	Delhi	15-04-2017	Delhi Daredevils	Kings XI Punjab	Delhi Dare
16	2017	Mumbai	16-04-2017	Gujarat Lions	Mumbai Indians	Mumbai In
17	2017	Bangalore	16-04-2017	Rising Pune Supergiants	Royal Challengers Bangalore	Royal Cha
18	2017	Delhi	17-04-2017	Delhi Daredevils	Kolkata Knight Riders	Delhi Dare
19	2017	Hyderabad	17-04-2017	Sunrisers Hyderabad	Kings XI Punjab	Kings XI P
20	2017	Rajkot	18-04-2017	Royal Challengers Bangalore	Gujarat Lions	Gujarat Lic
21	2017	Hyderabad	19-04-2017	Sunrisers Hyderabad	Delhi Daredevils	Sunrisers

Showing 1 to 22 of 636 entries

Data
ipldata 636 obs. of 18 variables

```
> ipldata<-read.csv("matches.csv")
> View(ipldata)
>
```

Fig.2. Loading of the Dataset

The importance of extractor function is variable importance measure produced by random forest.

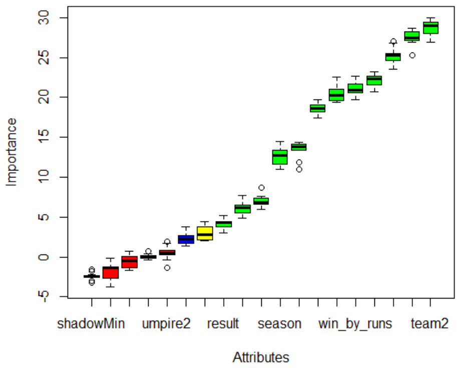


Fig.3. Results of Feature Selection using Boruta

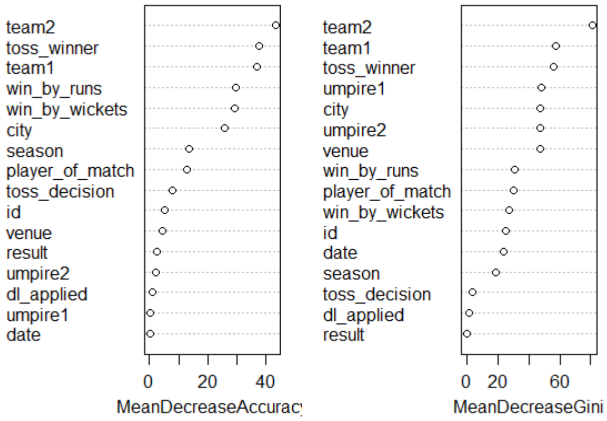


Fig.4. Results of Feature Selection using importance

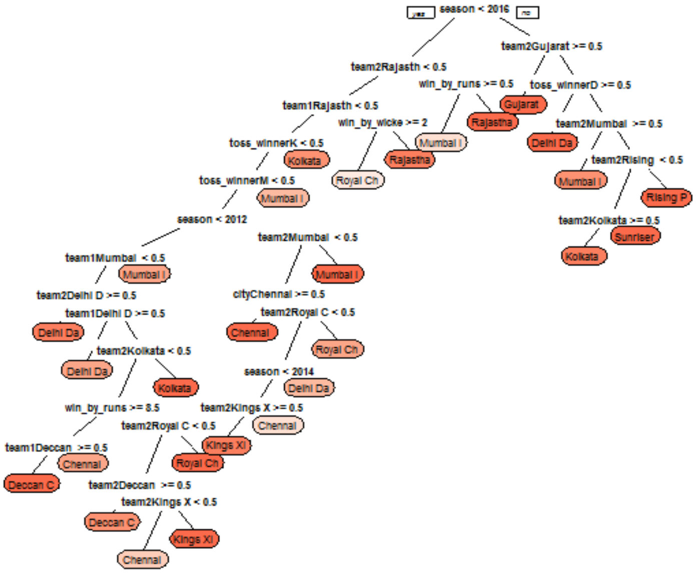


Fig.5. Decision Tree Results

Table.2. Decision Tree – Confusion Matrix

	CSK	DC	DD	GL	KXIP	KTK	KKR	MI	PW	RR	RPS	RCB	SRH
CSK	7	0	1	0	2	0	0	2	1	0	0	0	2
DC	0	2	0	0	0	0	0	0	0	0	0	0	0
DD	0	1	8	0	3	0	0	1	0	0	0	0	1
GL	0	0	0	2	0	0	0	0	0	0	0	0	1

KXIP	0	1	0	0	6	0	0	0	0	0	0	0	0
KTK	0	0	0	0	0	0	0	0	0	0	0	0	0
KKR	3	1	1	0	1	0	9	1	0	0	0	0	1
MI	3	0	1	0	1	1	0	12	1	0	0	0	1
PW	0	0	0	0	0	0	0	0	0	0	0	0	0
RR	0	0	0	0	0	0	0	0	0	12	0	0	0
RPS	0	0	1	0	0	0	0	0	0	0	7	1	1
RCB	2	0	0	0	0	0	4	0	0	0	1	0	6
SRH	0	0	0	0	0	0	1	1	0	0	1	1	1

Table.3. Random Forest - Confusion Matrix

	CSK	**DC**	**DD**	**GL**	**KXIP**	**KTK**	**KKR**	**MI**	**PW**	**RR**	**RPS**	**RCB**	**SRH**
CSK	50	0	0	0	1	0	2	4	0	3	0	0	0
DC	3	9	1	0	1	0	1	0	0	2	0	1	0
DD	2	1	28	0	3	0	0	2	0	3	0	3	0
GL	0	0	1	6	1	0	0	0	0	0	0	1	0
KXIP	4	0	2	0	31	0	5	4	0	2	0	2	2
KTK	0	0	0	0	0	3	2	0	0	1	0	0	0
KKR	1	0	2	0	3	0	50	4	0	0	0	4	1
MI	1	0	0	0	2	0	1	53	0	2	0	2	2
PW	2	0	1	0	2	0	2	1	1	1	0	0	0
RR	1	0	1	0	3	0	2	0	0	37	0	1	0
RPS	0	0	0	0	0	0	2	1	0	0	7	1	1
RCB	4	1	1	0	3	0	0	1	0	2	1	44	0
SRH	1	0	1	0	1	0	2	2	0	0	1	0	26

Table 4. Naive Bayes – Confusion Matrix

	CSK	**DC**	**DD**	**GL**	**KXIP**	**KTK**	**KKR**	**MI**	**PW**	**RR**	**RPS**	**RCB**	**SRH**
CSK	51	1	2	0	4	0	3	3	1	1	0	3	0
DC	1	24	3	0	2	0	2	3	1	0	0	3	0
DD	4	2	37	1	2	0	4	1	1	2	0	6	1
GL	0	0	3	11	2	0	0	2	0	0	0	2	0
KXIP	4	1	2	0	42	0	3	5	0	3	0	3	4
KTK	0	0	0	0	0	6	0	0	0	0	0	0	0
KKR	3	0	2	0	3	0	54	4	0	1	0	4	1
MI	5	0	2	0	5	0	1	55	1	3	2	6	1
PW	2	0	2	0	1	0	1	1	8	1	0	0	0
RR	3	0	2	0	1	0	1	3	0	46	0	3	2
RPS	0	0	1	0	0	0	0	1	0	0	10	1	0
RCB	2	1	3	0	6	0	3	6	0	3	1	38	2
SRH	1	0	2	0	1	0	2	2	0	2	1	1	29

For the four Classification algorithms, the dataset has been subdivided into training and testing subset. With the training subset, a model has been built. With the help of the model built, the test data is tested and the values are predicted and the Confusion matrices are formed.

Based on the testing data results, the above confusion matrices are formed with the predict() function. Analysing the past ten year records of the IPL results, CSK has won two times, KKR has won two times and MI has won three times. But when considering the accuracy and predicting the results, the overall winning pattern has to be considered. Giving the most importance to the three teams, the accuracy is derived with the help of the Confusion Matrices - True Positive, False Positive, True Negative and False Negative values.

From the Visualization of Decision Tree, it is inferred that, KKR has the more chances of winning based on the past ten year history. With the help of the Confusion matrices, all the four algorithms give the result as KKR which the more probability of winning.

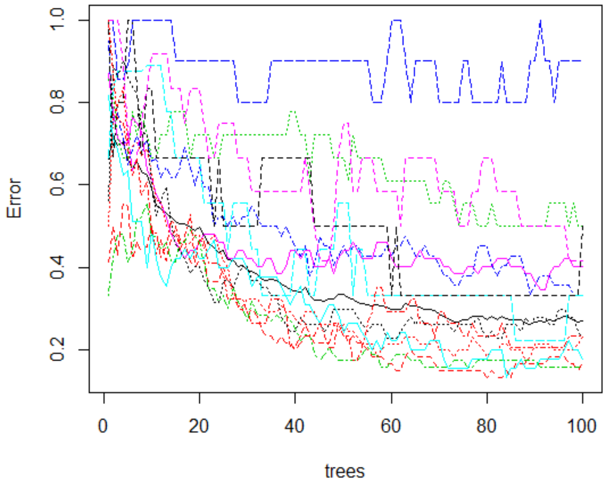


Fig.6. Random Forest Result

Based on the above results, we can conclude that Random Forest is performing well than the other algorithms as the measures are outstanding. It is with high accuracy and less error rate.

5. CONCLUSION

This work aims at understanding the dataset of past 10 years history of the IPL data. It helps to understand the four different machine learning algorithms working principal and their implementation in R. It creates the Model and Training dataset and helps to predict with the help of the model created. The model classifies the data and compares the results. It takes into consideration the measures accuracy, error rate, precision, recall, sensitivity and specificity. Based on this the best algorithm is selected as Random Forest. This work focuses on exploring IPL data and presenting its insights as graphical representation and comparative analysis. By making use of this, Indian Premier League and the fan followers can take decisions on the team's performance and predict the trophy winners that will lead to success in future.

REFERENCES

[1] S. Abhishek, Ketaki V. Patil, P. Yuktha and S. Meghana, Predictive Analysis of IPL Match Winner using Machine Learning Techniques", *International Journal of Innovative Technology and Exploring Engineering*, Vol. 9, No. 1, pp. 430-435, 2019.

[2] Sanjay Gupta, Hitesh Jain, Asmit Gupta and Hemant Soni, "Fantasy League Team Prediction", *International Journal of Research in Science and Engineering*, Vol. 6, No. 3, pp. 97-103, 2017.

[3] Pabitra Kumar Dey, Gangotri Chakraborty, Purnendu Ruj and Suvobrata Sarkar, "A Data Mining Approach on Cluster Analysis of IPL", *International Journal of Machine Learning and Computing*, Vol. 2, No. 4, pp. 351-354, 2012.

[4] Raza Ul Mustafa, M. Saqib Nawaz, M. Ikram Ullah Lali, Tehseen Zia and Waqar Mehmood, "Predicting the Cricket Match outcome using Crowd Opinions on Social Networks: A Comparative Study of Machine Learning Methods", *Malaysian Journal of Computer Science*, Vol. 30, No. 1, pp. 63-76, 2017.

[5] Rameshwari Lokhande and P.M. Chawan, "Live Cricket Score and Winning Prediction", *International Journal of Trend in Research and Development*, Vol. 5, No. 1, pp. 30-32, 2018.

[6] Vignesh Veppur Sankaranarayanan and Junaed Sattar,"Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction", *Proceedings of SIAM Conference on Data Mining*, pp. 1-7, 2014.

[7] Parag Shah, "Predicting Outcome of Live Cricket Match using Duckworth-Lewis Par Score", *International Journal of Latest Technology in Engineering, Management and Applied Science*, Vol. 6, No. 7, pp. 72-75, 2017.

[8] Amal Kaluarachchi and S. Aparna, "A Classification based Tool to Predict the outcome in ODI Cricket", *Proceedings of 5th International Conference on Information and Automation for Sustainability*, pp. 233-237, 2010.

[9] C. Deep Prakash Dayalbagh, C. Patvardhan and C. Vasantha Lakshmi, "Data Analytics based Deep Mayo Predictor for IPL-9", *International Journal of Computer Applications*, Vol. 152, No. 6, pp. 6-11, 2016.

[10] Jayshree Hajgude, Aishwarya Parameshwaran, Krishna Nambi, Anupama Sakhalkar and Darshil Sanghvi, "IPL Dream Team-A Prediction Software Based on Data Mining and Statistical Analysis", *International Journal of Computer Engineering and Applications*, Vol. 9, No. 4, pp. 113-119, 2015.

[11] Sonu Kumar and Sneha Roy, "Score Prediction and Player Classification Model in the Game of Cricket using Machine Learning", *International Journal of Scientific and Engineering Research*, Vol. 9, No. 2, pp. 237-242, 2018.