



International Journal of Physical Education, Sports and Health

P-ISSN: 2394-1685
E-ISSN: 2394-1693
Impact Factor (RJIIF): 5.38
IJPESH 2024; 11(5): 400-403
© 2024 IJPESH
<https://www.kheljournal.com>
Received: 21-08-2024
Accepted: 26-09-2024

Dr. B Navaneethan
M.P.Ed., MPhil., Ph.D.,
Director of Physical Education,
PSG College of Arts & Science,
Coimbatore, Tamil Nadu, India

Comparison of machine learning models for prediction the IPL match outcomes

Dr. B Navaneethan

Abstract

This paper presents a method aimed at predicting the outcome of the IPL matches by implementing machine learning algorithms. The proposed model consists of statistical data from the Indian Premier League matches which has been collected from trusted sports websites. Machine learning algorithms such as linear regression, decision tree regressor, random forest, XG Boost has been used to predict the match results, and their performance was compared using metrics such as accuracy, precision, recall and F_1 score. The performance metrics of the train and test predictions has been compared to evaluate the overfitting and underfitting models. The model that is neither overfitting or underfitting the unseen data and with higher performance metrics shall be chosen to predict the future IPL match outcomes. To implement the proposed model, the data is preprocessed into numerical values to implement the algorithms. The experimental setup demonstrates that the model gives up to 91.35% accuracy.

Keywords: IPL T20, machine learning algorithms and regression

Introduction

IPL T20 is one of the most famous and most-watched cricket series. T20 Cricket was first introduced in 2003 by England. T20 matches gained popularity because of their shorter format. BCCI initiated Indian Premier League (IPL) T20 in 2008. This tournament is organized every year. The use of statistical and machine learning models to analyze the various features of cricket including individual match results prediction has gained its importance. Due its popularity and the volume of money involved in these tournaments, the algorithm that accurately predicts the result of the matches has gained huge need and demand. Prediction of the outcome of a match using machine learning algorithms is performed using the records of the past performance of players and other related data. The data can be analyzed to create models that predict the winning team. This model can be created using the machine learning algorithms such as Decision Tree, Random forest, logistic regression and XG Boost and their results can be compared based on the Evaluation Measures as accuracy, precision, recall, sensitivity and error rate. The rest of the paper is organized as follows: Section II represents the related works done in the field of player's performance prediction in different sports using various machine learning methods, Section III describes data collection and the key features selection, Section IV presents algorithms and techniques, Section V presents the experimental results and analysis and finally in Section VI the conclusion has been drawn including future ideas.

Background Study

Gupta *et al.* [2] determined that the selection of the best team is always required by the management for best outcome. The research article emphasises the optimal solution to select the best team using Data Mining Techniques as an alternate to traditional method which is time consuming and cumbersome. Selecting the best is the most important task when a tournament is announced to win the same. In [3], the authors implement the fuzzy clustering logic. Clustering method was used to group the players based on IPL batting Statistics and it gave efficient and effective with higher accuracy rate. Raza Ul Mustafa *et al.* presented a study [4] on the finding of the feasibility of using the Twitter data to forecast the results of the match. The work has been proposed to use the machine learning techniques' robustness on the data

Corresponding Author:
Dr. B Navaneethan
M.P.Ed., MPhil., Ph.D.,
Director of Physical Education,
PSG College of Arts & Science,
Coimbatore, Tamil Nadu, India

collected from social media networks and other real world events to predict the match results. Support Vector Machine, Naive Bayes Classifier and the Linear Regression were used for this effort. Live Cricket Score and Winning Prediction work ^[5] explains about the building of the machine learning model that can predict the score for the chasing team and will estimate the score of the second innings of match. The proposed work uses the 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.

Data and Tools

The data for the study spanned between 2008 to 2023 and are

collected from espnocricket.com ^[11] which is the recognized and authenticated source for cricket data. The data were extracted from the website and Microsoft excel has been used to do the initial curation of the data. Later the data has been analysed using Python 3.8 open source software.

Problem Definition

Using the past sixteen years' data of IPLT20, the objective of this article is to predict the outcome fo the match held between two teams in the IPLT20 tournament at a specific venue using various Machine Learning algorithms. The data collected will be preprocessed and analyzed using various statistical tools and techniques. The preprocessed data will be used to train different classification based machine learning models in order to predict the future outcomes.

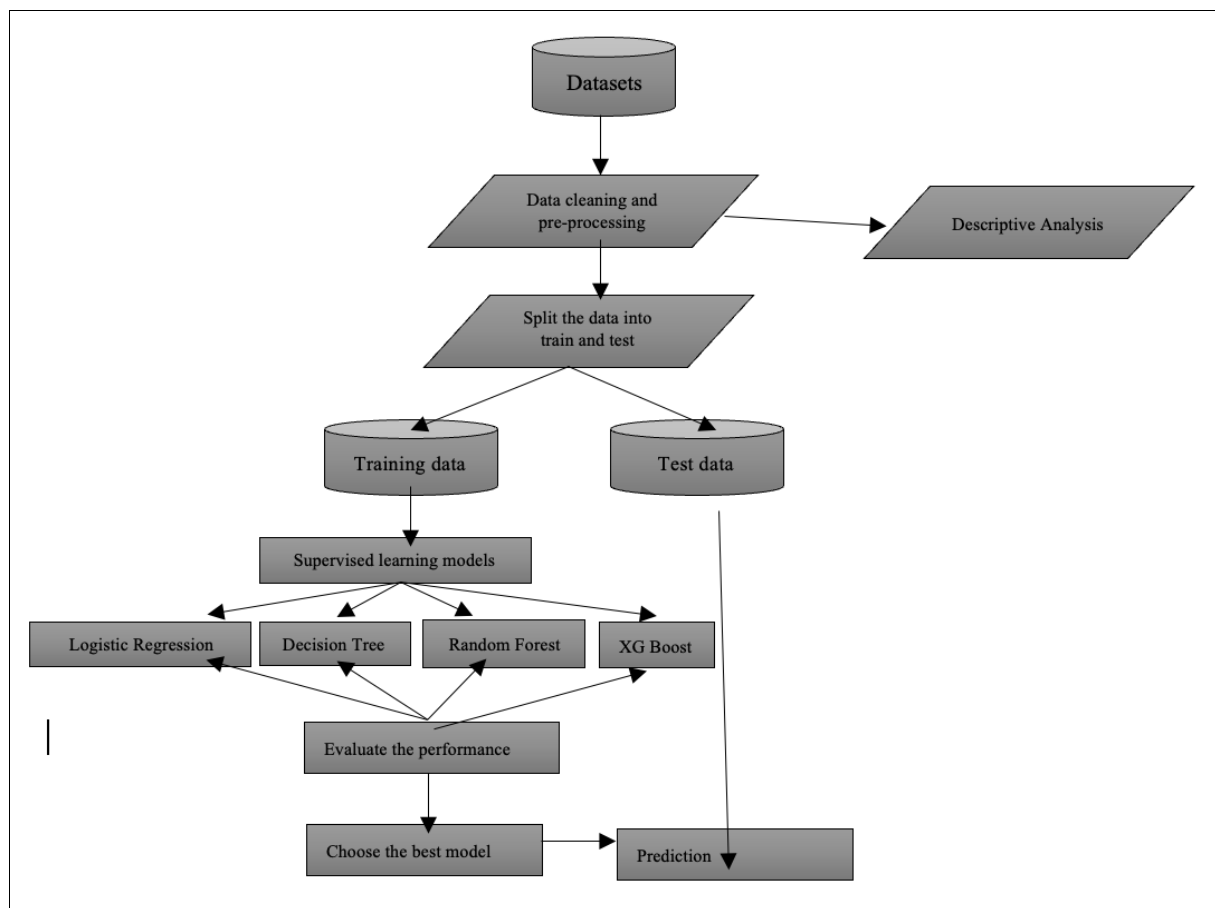


Fig 1: Model Architecture

Datasets

The Player's dataset contains the columns, namely Team 1, Team 2, Ground, WinningTeam, Match Date, and Margin. This dataset tells about the teams played in the particular match, who won the match, what was the margin for winning the match, the place where the match we conducted.

Feature Engineering

One hot encoding has been used to generate the columns for Team 1 and Team 2. This method helps to analyse the combination of the participating teams against the target variable "Winner", who has won the match in the given combination.

Splitting the data

The dataset was splitted into training and test data. The

training data was used to train the Machine learning models. The test data was used on the trained model to predict the target variables. The data from 2008 to 2022 was taken as the training data and data for the year 2023 was taken as the test data.

Training the data

The training data was fed into the supervised learning models to train the models based on the historical data. The trained models were evaluated based on the metrics such as Accuracy, Precision, Recall, and F_1 score.

Test data Prediction

The trained models were used on the unseen test data for prediction. The model were evaluated for over fitting / underfitting by comparing the metrics such as Accuracy,

Precision, Recall, and F_1 score of training data predicted and the test data predicted.

Supervised Learning Models: Supervised learning models such as logistic regression, decision tree classifier, random forest classifier and XG Boost have been used as the prediction models. GridsearchCV has been used for hyper parameter tuning and to find the best parameters for decision tree and random forest based on accuracy.

Experimental Results

Analysis of player dataset

Descriptive statistics and visualization methods have been used to understand the distribution of the data and to find the teams that has won the most of the matches and the teams that have lost the most of the matches. Table 1 gives the probability of team 1 winning against the team 2 in a match. The probability greater than 80% has been highlighted in the table.

Table 1: Probability of Team 1 Winning Against Team 2

	CSK	Chargers	DC	Daredevils	GT	Guj Lions	KKR	Kings XI	Kochi	LSG	MI	Punjab Kings	RCB	RR	SRH	Supergiant	Warriors
CSK		57%	67%	73%	50%		69%	69%	100%	100%	38%	0%	75%	73%	73%		67%
Chargers	33%			29%			33%	17%			33%		67%	33%			50%
DC	40%				0%		67%	50%		0%	29%	67%	60%	75%	50%		
Daredevils	43%	50%				50%	50%	50%	0%		75%		20%	50%	33%	100%	33%
GT	67%		50%				0%			100%	67%	0%	100%	67%	100%		
Guj Lions				0%			50%	0%			50%		50%		0%	100%	
KKR	42%	100%	50%	78%	0%	0%		62%	0%	0%	29%	67%	64%	67%	71%	0%	50%
Kings XI	50%	50%	100%	70%		0%	29%				43%		58%	50%	50%	100%	67%
Kochi	100%	0%		0%			100%	0%					0%	100%			
LSG	50%		100%		0%		100%				67%	0%	0%	0%	100%		
MI	47%	50%	50%	80%	100%	50%	73%	50%	0%	0%		33%	60%	67%	55%	33%	67%
Punjab Kings	50%		0%		0%		50%			0%	50%		67%	0%	0%		
RCB	40%	60%	50%	45%	50%	67%	50%	50%	100%	0%	41%	50%		43%	45%	0%	100%
RR	69%	100%	40%	75%	0%		67%	67%	100%	50%	58%	0%	40%		44%		100%
SRH	25%		20%	67%	100%	100%	45%	88%		0%	50%	33%	45%	44%		0%	100%
Supergiant				100%		100%	0%	200%			100%		100%		100%		
Warriors	33%	0%		33%			0%	67%	100%		0%		0%	50%	0%		

Model Performance: After the data pre-processing the primary challenge is to identify a machine learning algorithm that is best suited for predicting the target variable in the dataset. The target variable proposed in this article is the

outcome of the tournament. Since the output is a categorical variable, classification algorithms have been used. Table 2 gives the comparison of performance metrics between train and test data.

Table 2: Model Performance of Classification Models

	Train Data				Test Data			
	Accuracy	Precision	Recall	F_1 - Score	Accuracy	Precision	Recall	F_1 - Score
Logistic Regression	88.20	86.10	90.30	88.17	86.10	85.05	87.15	86.09
Decision Tree	90.30	89.25	92.40	90.81	89.25	86.10	90.30	88.17
Random Forest	93.45	89.25	93.45	91.33	90.30	88.20	91.35	89.76
XG Boost	93.45	89.25	93.45	91.33	91.35	89.25	92.40	90.81

From the above table, it is observed that XG Boost model is the best-performing model for predicting the outcome of the forth coming match.

Conclusion

In the recent days machine learning model is being utilized in almost all the real-life scenarios. Hence, using machine learning models in the game of cricket we can accurately predict the outcome and performance of the team and the players. Predicting the outcome of the tournament facilitates the team to further analyse the strategies and team selection and paying orders based on the opponents and the venue. The models can be used for predicting the performance of the individual players of the team using their historical data. The scope of such analysis can be extended across different sports and games in future research.

References

1. Abhishek S, Patil KV, Yuktha P, Meghana S. Predictive Analysis of IPL Match Winner using Machine Learning Techniques. *Int. J Innov. Technol. Explor. Eng.* 2019;9(1):430-435.
2. Gupta S, Jain H, Gupta A, Soni H. Fantasy League Team Prediction. *Int. J Res. Sci. Eng.* 2017;6(3):97-103.
3. Dey PK, Chakraborty G, Ruj P, Sarkar S. A Data Mining Approach on Cluster Analysis of IPL. *Int. J Mach Learn Comput.* 2012;2(4):351-354.
4. Mustafa RU, Nawaz MS, Lali MIU, Zia T, Mehmood W. Predicting the Cricket Match outcome using Crowd Opinions on Social Networks: A Comparative Study of Machine Learning Methods. *Malays J Comput. Sci.* 2017;30(1):63-76.
5. Lokhande R, Chawan PM. Live Cricket Score and Winning Prediction. *Int. J Trend Res Dev.* 2018;5(1):30-32.
6. Sankaranarayanan VV, Sattar J. Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction. In: *Proceedings of SIAM Conference on Data Mining*; 2014. p. 1-7.
7. Shah P. Predicting Outcome of Live Cricket Match using Duckworth-Lewis Par Score. *Int. J Latest Technol. Eng. Manag. Appl. Sci.* 2017;6(7):72-75.
8. Lemmer HH. The combined bowling rate as a measure of bowling performance in cricket. *S Afr. J Res. Sport Phys.*

- Educ Recreat. 2002;24(2):37-44. Available from: <https://hdl.handle.net/10520/EJC108746>
9. Bhattacharjee D, Pahinkar DG. Analysis of performance of bowlers using combined bowling rate. Int. J Sports Sci. Eng. 2012;6(3):184-192.
 10. Mukherjee S. Quantifying individual performance in Cricket-A network analysis of Batsmen and Bowlers. Physica A Stat Mech Appl. 2014;393:624-637. Available from: <https://doi.org/10.1016/j.physa.2013.09.027>
 11. ESPNcricinfo. Cricket Teams, Scores, Stats, News, Fixtures, Results, Tables. Available from: <http://www.espnricinfo.com/>
 12. ESPNcricinfo. Cricket Teams, Scores, Stats, News, Fixtures, Results, Tables. Available from: <https://tnpl.cricket/stats/>