International Journal of Physical Education, Sports and Health 2024; 11(5): 392-397

P-ISSN: 2394-1685
E-ISSN: 2394-1693
Impact Factor (RJIF): 5.38
IJPESH 2024; 11(5): 392-397

https://www.kheljournal.com
Received: 16-08-2024
Accepted: 20-09-2024

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

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

Player's performance prediction in TNPL T20 using machine learning algorithms

Dr. B Navaneethan

Abstract
This paper presents a method aimed at predicting a cricket player's upcoming match performance by implementing machine learning algorithms. The proposed model consists of statistical data of players of Tamil Nadu Premier League cricket team which has been collected from trusted sports websites, feature selection algorithms univariate selection, and machine learning algorithms such as linear regression, decision tree regressor, random forest, and support vector machine with linear and polynomial kernel. To implement the proposed model, the data is preprocessed into numerical values to implement the algorithms. Furthermore, feature selection algorithms are applied to extract the attributes that are highly correlated to the output feature. The machine learning algorithms are used to predict runs scored by a batsman in the upcoming match. The experimental setup demonstrates that the model gives up to 92.35% accuracy. Therefore, this will help calculate the player's future performance and thus ensure better team selection for forthcoming cricket matches.

Keywords: Player's performance prediction, cricket matches, cricket's origins, sports websites

Introduction
Cricket's origins can be traced back to Surrey in South East England in the 16th century, when this "bat and ball" sport was mentioned. Cricket is all about numbers: How many runs a batsman scores, how many wickets a bowler takes, how many games a team wins, how often a batter reacts a certain way to a bowling attack, etc. With the help of robust analytics tools and numerical computing software like Numbly, it is possible to examine cricket statistics for the purpose of enhancing performance as well as researching company prospects, the market as a whole, and cricket economics. Analytics for cricket offer insightful information about the game and forecast game results.

These days, there are a wealth of nearly limitless resources for cricket game records and statistics, such as ESPN Cricinfo and Cricsheet. Using the most recent machine learning and predictive modelling methods, these and other cricket databases have been used for cricket analysis. Technology and analytics are used by media and entertainment outlets as well as professional sports organisations involved in the game to identify critical indicators that increase the likelihood of winning a match such as moving average of batting performance, score prediction, understanding a player's fitness and effectiveness against various opponents, and calculating a player's contribution to wins and losses to inform tactical choices about the makeup of the team.

The provision of a huge quantity of pertinent data has emerged as the most important aspect of the subject of machine learning, which seeks to concentrate on more intricate and large-scale tasks [1]. This proposed model's primary goal is to determine the players' performance from the player pool by looking at their historical records. The goal of the suggested model is to forecast a batsman's run that he will score in the upcoming game. The method involves applying various supervised learning techniques to compare the accuracy of the models.

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

The performance of cricket players has been the studied in very few research studies. Muthuswamy and Lam [3] graphed the effectiveness of Indian bowlers against seven foreign teams frequently confronted by the Indian cricket team. They used the back propagation system and the radial basis network function to predict a bowler's potential runs conceded, and wickets taken in an ODI match. Wickramasinghe [4] used the hierarchical linear model to predict and forecast the batsman's performance in the test series. In limited-overs cricket, Barr and Kantor [5] calculated and selected the batsman based on a 2D graphical representation with strike rate on x-axis and P(out) on the y-axis, where, P(out) is the probability of the batsman getting out. The player selection criterion considered batting average, strike rate, and P(out). Iyer and Sharda [6] implemented neural networks to categorize bowlers and batsman into three clusters: performer, middling, and failure, considering players' historical ratings to recommend squad inclusions for the 2007 World Cup. Jhawar and Pudi [7] predicted the match results by analyzing individual player performances in playing teams. Using simulation algorithms, they simulated bowlers' bowling rate as a benchmark to evaluate bowlers' performances. Bhattacharjee and Pahinkar [9] evaluated IPL bowlers' performances by considering economy, strike rate, and bowling average, identifying other important variables affecting performance through a multiple regression model. Mukherjee [10] employed social network analysis to give ratings to the bowlers and batters in the team's performance, using data from ODI and test cricket match player statistics. Shah [11] introduced a new metrics for assessing the player performance. The new batting evaluation metrics incorporates the quality of each bowler a batsman encounters, while the bowlers' metrics consider the quality of each batter they bowl to. The overall performance index for a batsman is identified by finding the summation of individual performances against each bowler (batsman).

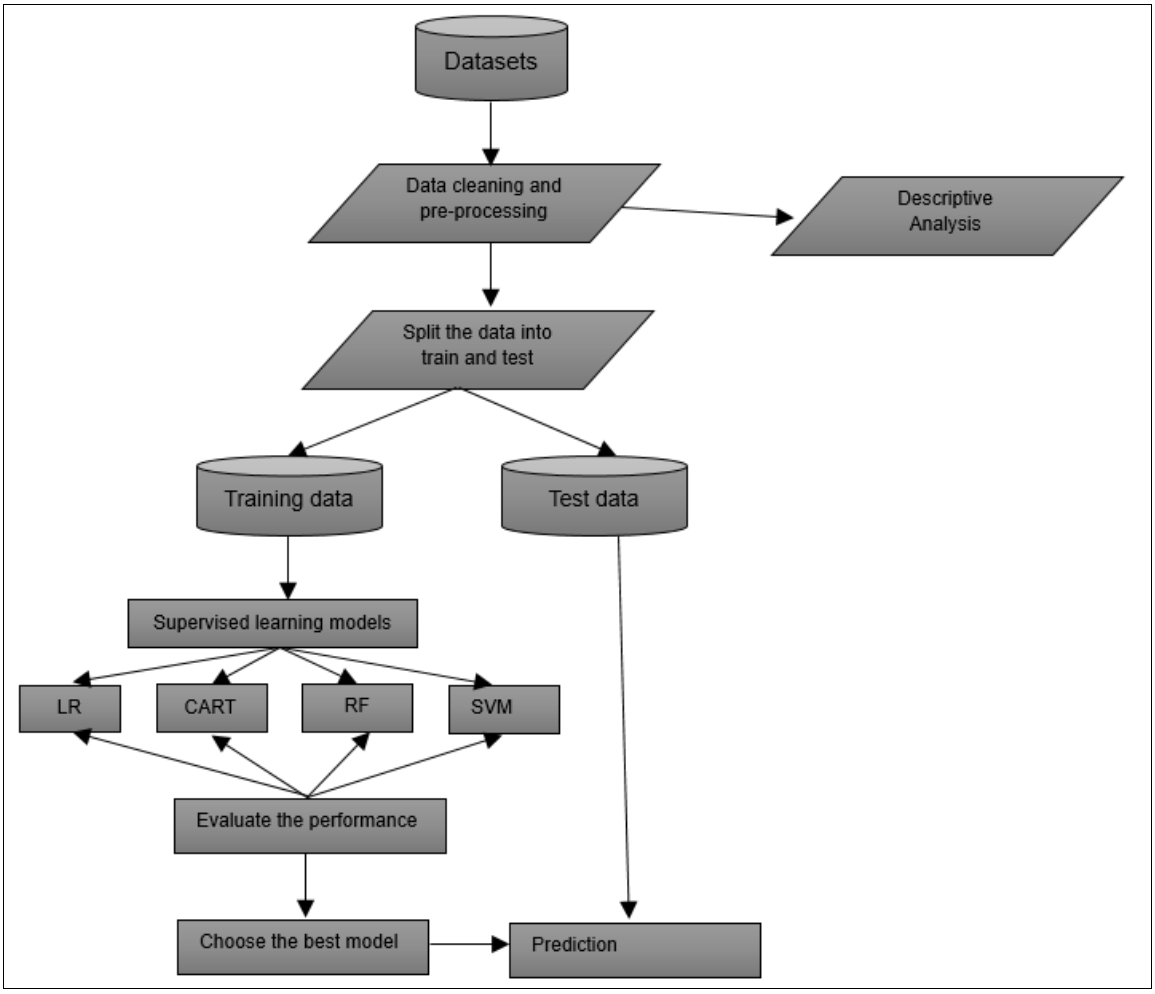
Data and Tools

The data for the study is collected from espncrickinfo.com [11] and https://tnpl.cricket [12] which are the recognized and authenticated sources for cricket data. The data is categorized by innings, i.e., each inning played by a player and their score. Some data attributes included are name, country, opposition, runs scored and the number of 4s and 6s. The objective is to have the most accurate data to analyze and train the model. The duplicate values are initially treated to eliminate using the data mining approach. The dataset with exactly 30 percent of the initial data was eliminated. Python 3.8 is used to analyse the data, training and fitting the model and evaluating the performance of the model.

Problem Definition

Using the past eight years' data of TNPL, the primary objective of this article is to predict the total runs scored by the batsman based on the analysis of historical data 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 regression based machine learning models in order to predict the future data.

Model Architecture



Datasets

The Player's dataset contains the columns, namely Player, Team, Balls, 4s, 6s, Opposition, StrikeRate, Ground, Match Date, Dismissed, and Runs. This dataset tells about the player and his performance against various teams.

Table 1: Shows the dataset column and its description.

S. No	Variable Name	Description
1	Player	Name of the player
2	Balls	Number of balls faced
3	4s	Number of 4's hit by the player
4	6s	Number of 6's hit by the player
5	Team	Team the player belongs to
6	Opposition	Opponent team against which the runs were scored
7	Ground	Ground in which the match was held
8	Match Date	Date on which the match was held
9	Dismissed	Where the player was out or not out
10	Runs	Total runs scored in the match
11	Year	The year in which the match was held

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 2016 to 2023 was taken as the training data and data for the year 2024 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 Mean Square Error, Mean Absolute Error and Mean Absolute Percentage Error.

Test data Prediction

The trained models were used on the unseen test data for prediction. The model were evaluated for over fitting / under fitting by comparing the metrics such as Mean Square Error, Mean Absolute Error and Mean Absolute Percentage Error of training data predicted and the test data predicted.

Supervised Learning Models

Supervised learning models such as linear regression, decision tree regressor, random forest regressor and support vector Machine have been used as the prediction models.

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 top 10 players in terms of the number of balls faced, number of 4s and 6s hit and number of runs scored. Chart 1 below gives the top 10 batsmen based on the total balls faced in the years between 2016 to 2024. Chart 2 below gives the top 10 batsmen based on the total runs scored in the years between 2016 to 2024. Chart 3 below gives the top 10 batsmen based on the total 4's scored in the years between 2016 to 2024. Chart 3 below gives the top 10 batsmen based on the total 4's scored in the years between 2016 to 2024.

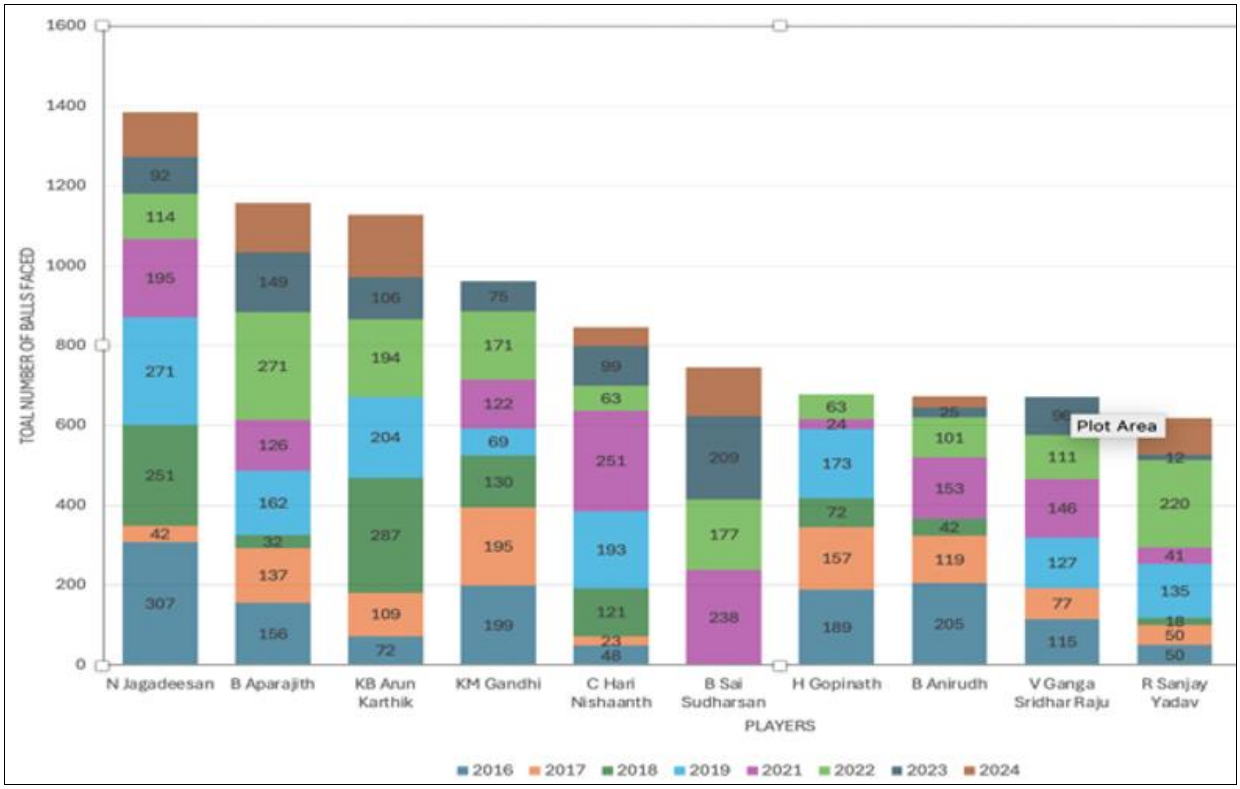


Chart 1: Top 10 Batsmen by total balls faced

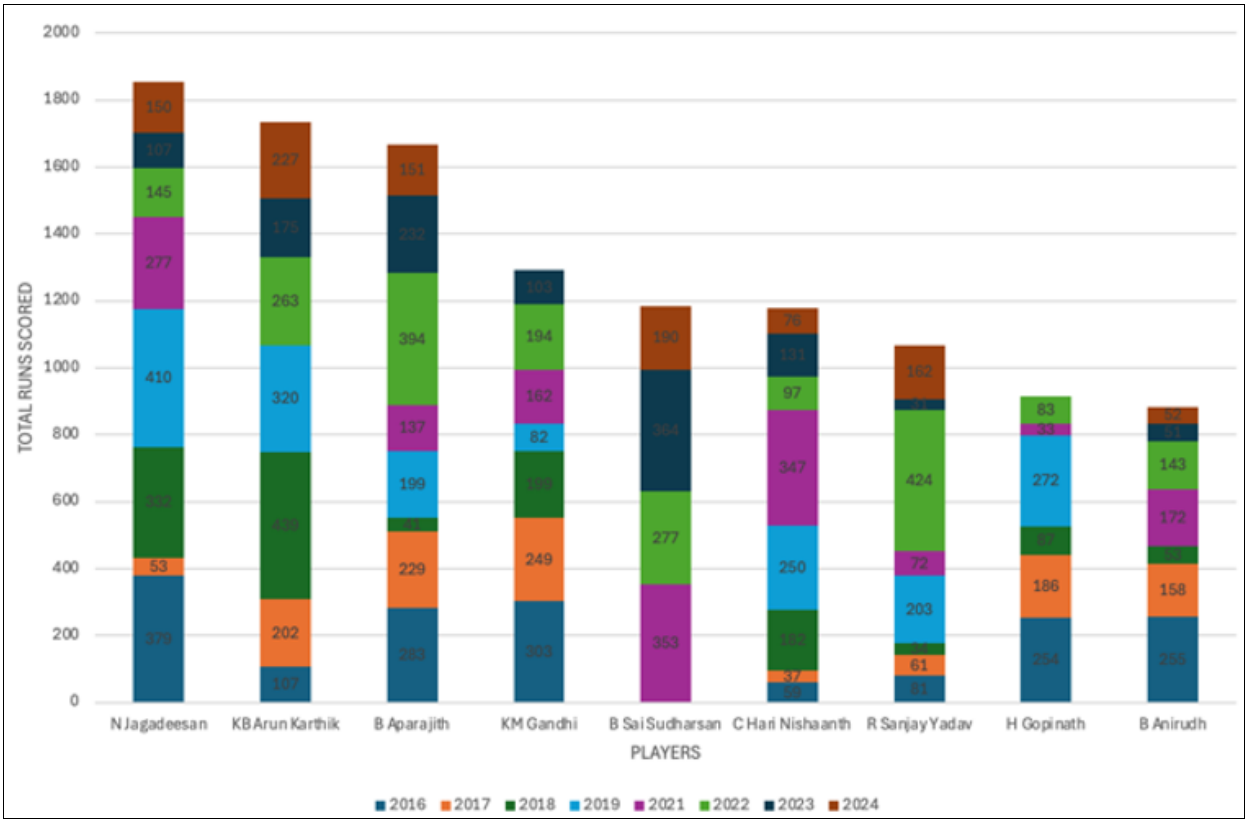


Chart 2: Top 10 Batsmen by total runs scored

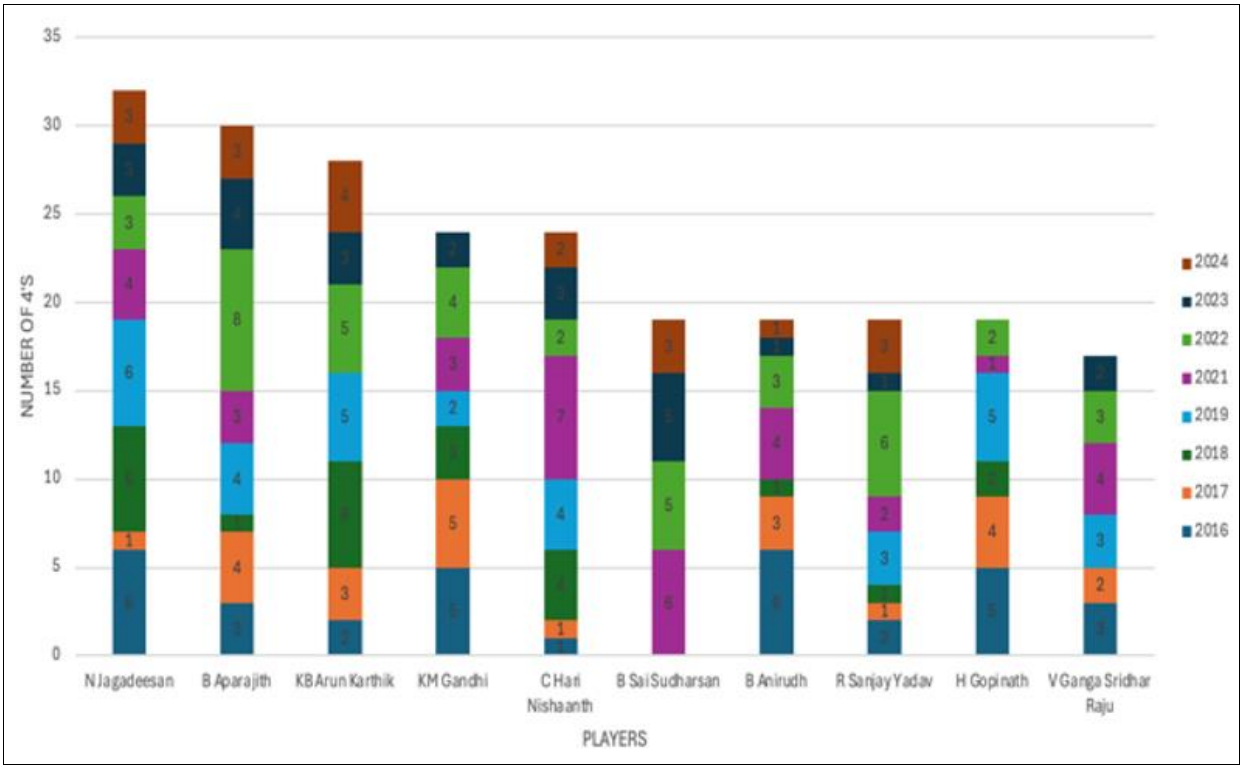


Chart 3: Top 10 players by number of 4s

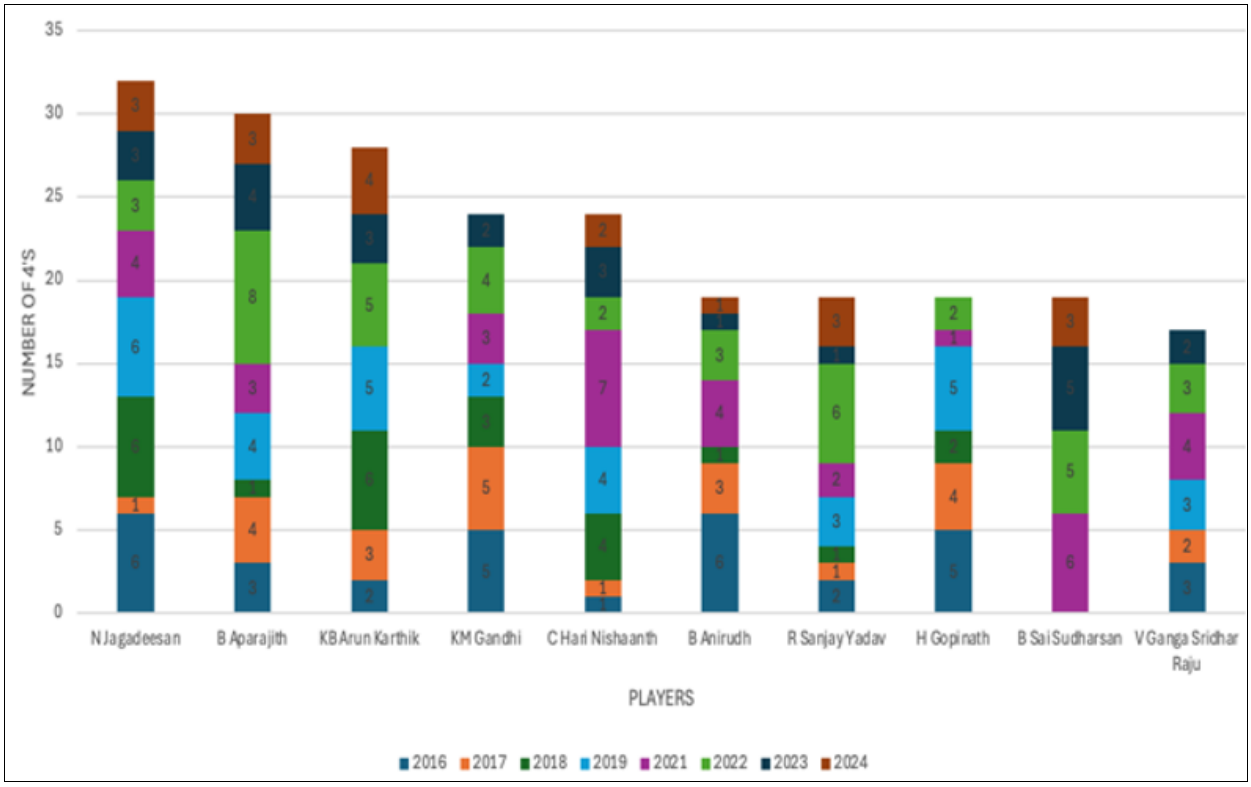


Chart 4: Top 10 players by number of 6s

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 run scored by the batsmen. Since the output is a continuous variable, regression algorithms have been used.

Table 2: Gives the comparison of performance metrics between train and test data

	Train Data			Test Data		
	MSE	**MAPE**	**MAE**	**MSE**	**MAPE**	**MAE**
Linear Regression	3.13	0.05	2.40	3.20	0.05	2.33
Decision Tree	2.68	0.04	1.86	5.24	0.04	3.64
Random Forest	2.15	0.23	1.17	3.65	0.23	1.87
Support Vector Machine	17.85	0.24	12.82	17.73	0.24	12.11

From the above table, it is observed that Random Forest Regressor model is the best-performing model for predicting the runs scored by the batsman in the fourth coming match.

Conclusion

In the recent day's 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 performance of the team and the players. Using these models the team will have an idea about the player's performance and the better team selection and management. In future, deep learning models can be used to predict the performance of the players and team that can give more accurate prediction than the matching learning models. The algorithms shall be used in the national level and international level performance evaluation also.

References

1. Ishi MS, Patil JB. A study on impact of team composition and optimal parameters required to predict result of cricket match. In: Social Networking and Computational Intelligence: Proceedings of SCI-2018. Singapore: Springer; c2020, p. 389-399. https://doi.org/10.1007/978-981-15-2071-6_32.
2. Sarangi S, Singh RR. Winning one-day international cricket matches: A cross-team perspective. J Bus Anal. 2023;6(1):39-58. https://doi.org/10.1080/2573234X.2022.2041370.
3. Muthuswamy S, Lam SS. Bowler performance prediction for one-day international cricket using neural networks. In: IIE Annual Conference Proceedings. Institute of Industrial and Systems Engineers (IISE); c2008, p. 1391-1395.
4. Wickramasinghe IP. Predicting the performance of batsmen in test cricket. J Hum Sport Exerc. 2014;9(4):744-751. https://doi.org/10.14198/jhse.2014.94.01.
5. Barr GDI, Kantor BS. A criterion for comparing and selecting batsmen in limited overs cricket. J Oper Res Soc. 2004;55(12):1266-1274. https://doi.org/10.1057/palgrave.jors.2601800.
6. Iyer SR, Sharda R. Prediction of athletes' performance using neural networks: An application in cricket team selection. Expert Syst Appl. 2009;36(3):5510-5522. https://doi.org/10.1016/j.eswa.2008.06.088.
7. Jhanwar MG, Pudi V. Predicting the outcome of ODI cricket matches: A team composition-based approach. MLSA@PKDD/ECML; c2016, 78.

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. 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. Phys A. 2014;393:624-637. https://doi.org/10.1016/j.physa.2013.09.027.
11. Cricket Teams, Scores, Stats, News, Fixtures, Results, Tables [Internet]. ESPNcricinfo. Available from: http://www.espncricinfo.com
12. Cricket Teams, Scores, Stats, News, Fixtures, Results, Tables [Internet]. ESPNcricinfo. Available from: https://tnpl.cricket/stats