



E-ISSN: 2707-7020
P-ISSN: 2707-7012
JSSN 2024; 5(2): 167-171
www.allsportsjournal.com
Received: 22-08-2024
Accepted: 27-09-2024

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

Using machine learning models for predicting the U-19 world cup (men) match outcomes

Dr. B Navaneethan

DOI: <https://doi.org/10.33545/27077012.2024.v5.i2c.288>

Abstract

This paper presents a method aimed at predicting the outcome of the U-19 WORLD CUP matches by implementing machine learning algorithms. The proposed model consists of statistical data from the U-19 World Cup 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 U-19 WORLD CUP 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 73.35% accuracy.

Keywords: U19 ODI, machine learning models, decision tree classification, random forest classification, logistic regression, XG boost

Introduction

National under-19 teams compete in the International Cricket Council's (ICC) ICC Under-19 Men's Cricket World Cup, an international cricket competition. It was first played as the Youth Cricket World Cup in 1988 and wasn't held again until 1998. Since then, the ICC has organized the World Cup as a biannual tournament. Although there were only eight teams in the tournament's inaugural edition, sixteen teams have participated in each of the tournament's subsequent iterations. India has won the World Cup five times, which is a record among the full members, compared to four victories for Australia, two for Pakistan, and one each for Bangladesh, England, South Africa, and the West Indies. Sri Lanka and New Zealand have advanced to the competition finals but lost. In cricket, predictive analytics uses machine learning algorithms and statistical models to estimate player performance and match results. This entails examining past player performance data, team records, and other match-related variables including weather and pitch conditions. 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

Kaluarachchi ^[1] classified the aspects that determine the outcome of the match, such as home game advantage, the impact of day/night on the toss, and batting first, using artificial intelligence techniques, notably Bayes classifiers in machine learning. The software tool known as CricAI is the end result of this work. Based on the supplied variables, such as whether home game advantage is available at the start of the match, the tool provides the likelihood of winning. When teams play cricket in the real world, the CricAI can be utilized. It is employed to alter specific elements in order to raise winning odds in the actual field.

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

The models make use of machine learning-related data analytics techniques. It could be challenging to anticipate the outcome of a match that is impacted by rain. Batting, bowling, fielding, team selection, result prediction, and target revision are all crucial in a rain-affected game. The mathematical model developed based on the knowledge of previous match results can be used to solve the match prediction. SVM is used in the creation of the predictor models. The Deep Mayo Predictor is used in this suggested work [2].

With the Bayesian Prediction Technique and parameter-based filtering, the statistical analysis and data mining software created based on Jayshree Hajgude's work [3] creates a dream team. The current IPL players' details are in the database. This work helps to extract the required data to be used by the prediction algorithm in order to predict the outcome of each player.

Using ODI data, a predictive algorithm is created to forecast cricket scores and individual performances [4]. Carried out supervised techniques like as Naïve Bayes and SVM. KNN and MLP clustering techniques are used for precise classification. In order to categorize the elements that impact the outcome of the match, such as home game advantage, the impact of day and night on the toss, and batting first, Kaluarachchi's study [1] employed artificial intelligence techniques, notably Bayes classifiers in machine learning. The software tool known as CricAI is the end result of this effort. Based on the supplied variables, such as whether home game advantage is available at the start of the match, the tool provides the likelihood of winning. It is employed to change specific elements in order to improve

the likelihood of success in real life.

The models utilize data analytics techniques from the domain of machine learning. Predicting the outcome of a match affected by rain can be quite challenging. In such matches, aspects like batting, bowling, fielding, team selection, result prediction, and target revision are crucial. The prediction of match outcomes can be addressed using a mathematical model developed from the insights gained from previous match results. Predictor models have been created with the assistance of Support Vector Machines (SVM). This approach has been introduced through the Deep Mayo Predictor [2].

Data and Tools

The study's data covers the period from 2000 to 2024, sourced from espncrickinfo.com, a reputable and verified cricket data source. The information was gathered from the website, and initial data curation was performed using Microsoft Excel. Subsequently, the analysis of the data was conducted using Python 3.8, an open-source software.

Problem Definition

For this article, the goal is to forecast the result of a match between two teams in the U-19 WORLD CUP ODI tournament at a designated venue, utilizing data from the past sixteen years. The gathered data will undergo preprocessing and will be examined through various statistical methods and tools. Once the data is processed, it will be employed to train multiple classification-based machine learning models aimed at predicting upcoming match outcomes.

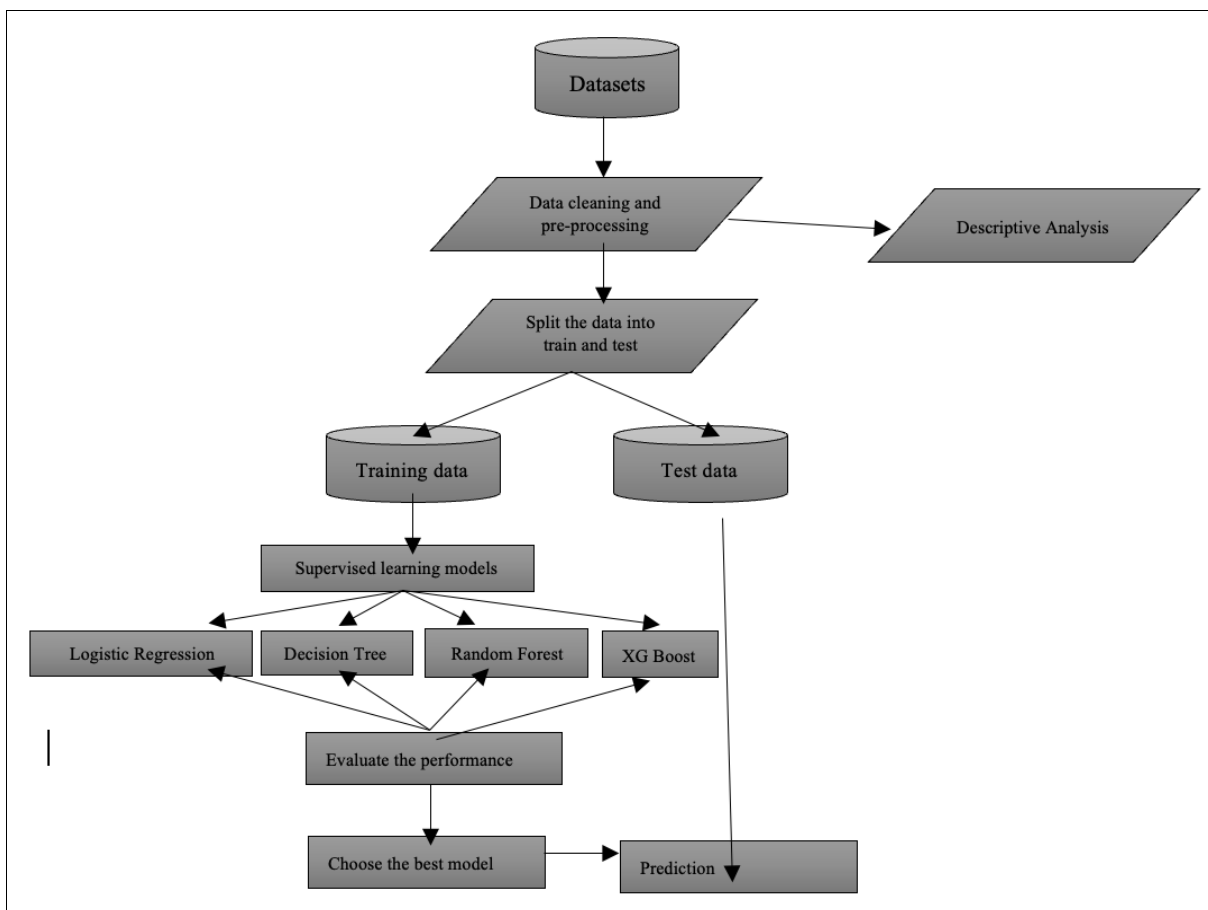


Fig 1: Model Architecture

Datasets

The Players dataset contains the columns Team 1, Team 2, Ground, Winning Team, Match Date, Margin. This dataset tells us which teams played in a particular match, who won the match, what was the margin by which they won the match, and the venue where the match was played.

Feature Engineering

One-hot encoding was used to generate the columns for Team 1 and Team 2. This method allows to analyze the combination of participating teams with respect to the target variable “Winner”, who won the match in the given combination.

Splitting the data

The data set was divided into training and testing data. The training data was used to teach machine learning models. The test data was used in the trained model to predict the target variable. Data from 2000 to 2023 was taken as training data and data from 2024 was taken as 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 F₁ 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 F₁ 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. Chart 1 shows the number of matches played by top 10 teams (number of matches) in the last five years between 2020 to 2024. Bangladesh U-19 team has played the most number of matches followed by England U-19 team and Afghan U-19 Team. India is in the 6th place.

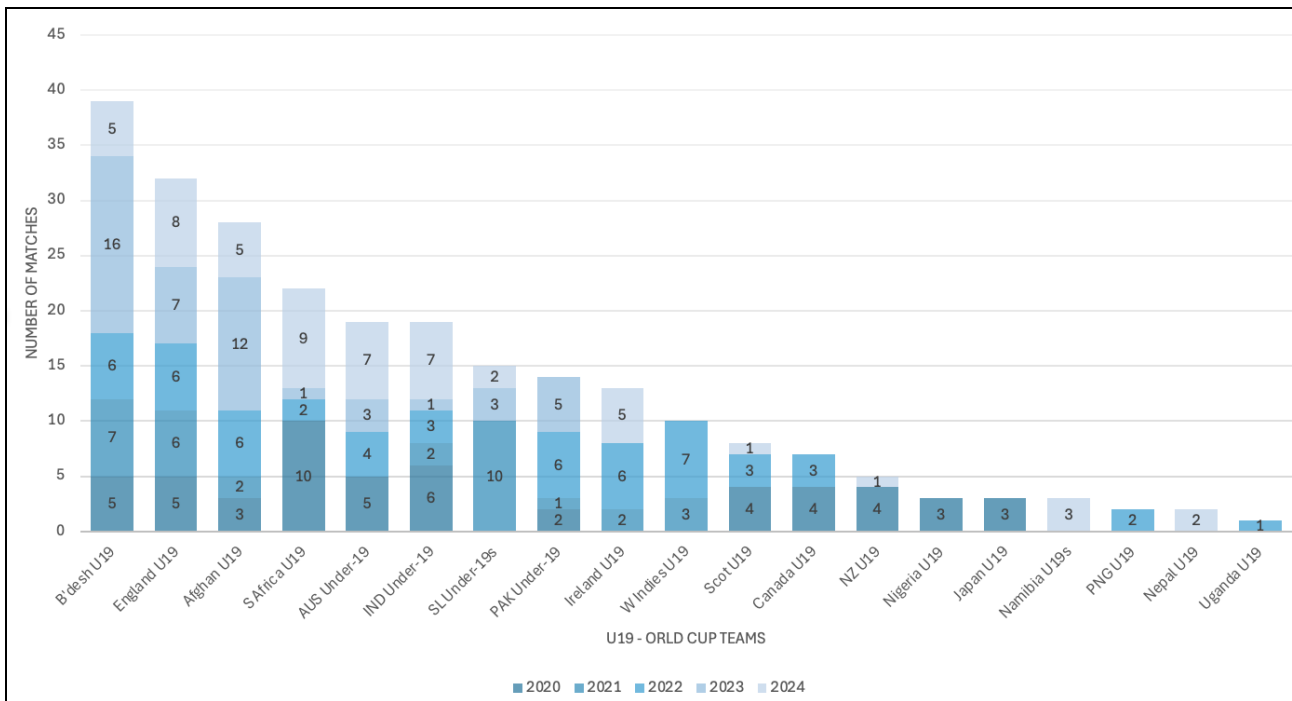


Chart 1: Number of matches played by top 10 teams between 2020-2024

Among the matches held between 2000 to 2024, India has won the number of matches (14 percent) followed by Bangladesh (12 Percent) and Pakistan (11 percent). Chart 2

represents the winning proportion of the teams between the period 2000 to 2024.

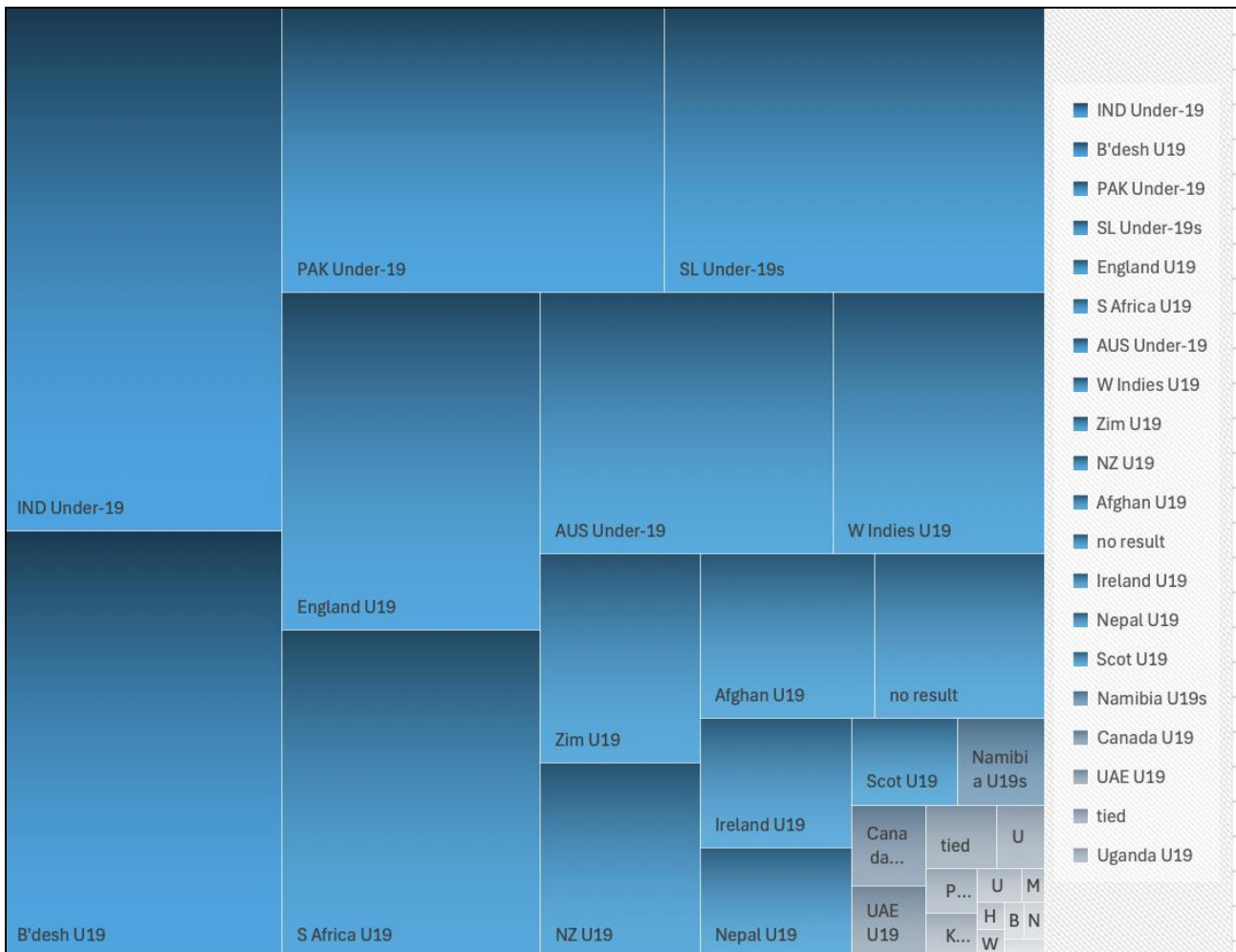


Chart 2: Proportion of Matches Won By the Teams (2000-2024)

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 1 gives the comparison of performance metrics between train and test data.

Table 1: Model Performance of Classification Models

	Train Data				Test Data			
	Accuracy	Precision	Recall	F ₁ - Score	Accuracy	Precision	Recall	F ₁ - Score
Logistic Regression	78.2	73.1	77.3	75.17	72.1	73.05	74.15	73.60
Decision Tree	72.3	79.25	81.4	80.32	77.25	69.1	79.3	74.02
Random Forest	73.45	72.25	80.45	76.24	70.3	68.2	72.35	70.24
XG Boost	79.45	72.25	77.45	74.80	73.35	71.25	75.4	73.30

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

Recently, machine learning models are used in almost all actual scenarios. Therefore, you can accurately predict the results and performance of teams and players using a machine learning model in cricket games. Predicting tournament outcomes allows teams to analyze their strategies in more detail as well as how team selection and payouts will play out based on opponents and location. The model can then be used to predict the performance of individual players on a team using historical data. In future studies, the scope of such analyses could be

extended to different sports and games.

References

1. Kaluarachchi A, Aparna S. A Classification based Tool to Predict the outcome in ODI Cricket. In: Proceedings of the 5th International Conference on Information and Automation for Sustainability. 2010. p. 233-237.
2. Dayalbagh CDP, Patvardhan C, Lakshmi CV. Data Analytics based Deep Mayo Predictor for IPL-9. Int. J Comput Appl. 2016;152(6):06-11.
3. Hajgude J, Parameshwaran A, Nambi K, Sakhalkar A, Sanghvi D. IPL Dream Team-A Prediction Software Based on Data Mining and Statistical Analysis. Int J Comput Eng Appl. 2015;9(4):113-119.
4. Kumar S, Roy S. Score Prediction and Player

- Classification Model in the Game of Cricket using Machine Learning. *Int. J Sci. Eng. Res.* 2018;9(2):237-242.
5. Bhattacharjee D, Pahinkar DG. Analysis of performance of bowlers using combined bowling rate. *Int J Sports Sci Eng.* 2012;6(3):184-192.
 6. 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.
 7. Lokhande R, Chawan PM. Live Cricket Score and Winning Prediction. *Int. J Trend Res Dev.* 2018;5(1):30-32.
 8. Sankaranarayanan VV, Sattar J. Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction. In: *Proceedings of the SIAM Conference on Data Mining.* 2014. p. 01-07.
 9. 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.
 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/>