Parametric analysis of a cricketer's performance using machine learning approach

Doddamani John Vestly
Department of Computer Science and Engineering,
Vardhaman College of Engineering,
Hyderabad, India
doddamanijohn123@gmail.com

Shanmugasundaram Hariharan
Department of Computer Science and Engineering,
Vardhaman College of Engineering,
Hyderabad, India
hariharan.s@vardhaman.org

Vinay Kukreja
Department of Computer Science and Engineering,
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
vinay.kukreja@chitkara.edu.in

Andraju Bhanu Prasad
Department of Computer Science and Engineering,
Vardhaman College of Engineering,
Hyderabad, India
a.bhanuprasad@vardhaman.org

Kondawar Swaraj
Department of Computer Science and Engineering,
Vardhaman College of Engineering,
Hyderabad, India
swarajkondawar@gmail.com

Devarakonda Gopichandh
Department of Computer Science and Engineering,
Vardhaman College of Engineering,
Hyderabad, India
dgchandhu7217@gmail.com

***Abstract*—This paper aimed to suggest the appropriate way to analyze the performance related to cricket players based on some predominant parameters such as batting, bowling, fielding, and overall contribution to the team success. In order to fetch the good result and success in cricket match ,it is important to select the best players for the match. This is considered to be one of the crucial decision successful match outcomes. Eventually right actions and right decisions leads to success. These actions and decisions can be made through perfect selection of the machine learning algorithms with correct data sets and accurate attributes as well. This study makes recommendations for the best data source, approximate properties, and machine learning techniques. This paper aims to identify the factors that contribute to good performance and to provide insights that can be used by coaches and selectors in identifying and developing talented players.**

Keywords— Performance; cricket; datasets; attributes; machine learning; algorithms

I. INTRODUCTION

Cricket is a sport that requires a diverse range of skills such as batting, bowling, fielding and teamwork. It is most popular sports in the world with a large number of players and fans. Basically it is a bat-and-ball game that involves two teams with each of 11 members, one team batting and other team is assigned to fielding. In order to win, the batting team must score as many runs as they can,while the fielding team's goal is to limit scoring and take wickets by removing the batsmen[1]. Cricket has a long and rich history, with its origins dating back to the 16th century in England. The game evolved from a simple bat-and-ball activity played by shepherds and farmers into a highly structured and organized sport that is now played at the international level. Today there are several different formats of the game that are played around the world [2].

The traditional form the game is known as Test cricket, which is played over five days and is considered the ultimate test of a player's skill and endurance. One-day cricket, also known as limited-overs cricket, is a shorter format of the game that is played over a single day and involves each team batting for a limited number of over's. In recent years, new format of the game called Twenty20 cricket had emerged which is even shorter and faster-paced than one-day cricket. In Twenty20 cricket, each team bats for a maximum of 20 overs and the matches are typically completed within three hours. This format of the game has become very popular particularly in India and other parts of Asia[3].

In addition to these main formats of the game, there are also many different domestic and international tournaments that are played throughout the year, including the ICC World Cup, the IPL(Indian Premier League) and the Australian BBL(Big Bash League).These tournaments attract huge crowds and generate a lot of excitement and interest among [4] cricket fans around the world. The use of of data science to evaluate cricket player's performance has gained popularity in recent years.This interest has been driven by the availability of large datasets on cricket matches, as well as advances in machine learning and deep learning tools as well. These tools have enabled researchers to identify patterns and trends in player performance, which can be used to improve team selection and coaching strategies [5].

II. RELATED WORK

Now-a-days machine learning took every aspect to be its enhanced version.At some point,[1] combined their two approaches to forecast cricket match results based on the performance of individual players. The first technique entails categorizing players, while the second method learns the correlation between player performance and match outcome using a shallow convolution neural network(CNN) with four layers. They developed a method for choosing the best player for sporting competitions or tournaments; it is important to evaluate their performance both nationally and internationally. With 80% accuracy, it described [2] how to use XG Boost and came to a conclusion.

The exploration of different aspects suggested a technique using the IPL data set. While calculating player performance, they [3] used studies that only took bowling and batting skills into account, leaving out skills like fielding and wicket keeping. A 34% accuracy breakthrough decision tree was attained by CatBoost. The research focuses on creating a prediction model to categorize players and assist with team projection in cricket, notably in the ODI format. They study took several algorithms for study like decision Tree (91.87%), random Forest(95.78%) and SVM Classifier(93.64%) which yielded results as illustrated respectively [4].

India's test team performance against every opposition to extract valuable information for future matches. They used several [5] data analytics and visualization techniques. The Random Forest Classifier model proved to be the most effective with an accuracy of 75%.The goal and examination of the study is to use machine learning to forecast the best player for a cricket match. Support Vector Machine, Naive Bayes, and Random Forest algorithms [6] were used to predict the team for Bangladesh's ODI cricket team with 94% accuracy for batters and 93%accuracy for bowlers.

In one another research study a method for selecting the most effective playing 11 for the Indian cricket team by taking into account a number of variables that have an impact on the player's performance is presented. The model offers a thorough understanding of a player's skill set and draws on data from reliable sources. It also focuses on [7] identifying the best all-around player and uses the Random Forest algorithm to reach 76% accuracy for batters, about 67% accuracy for bowlers, and 95% accuracy for all-rounders.

Neural networks and clustering algorithms can be used to select the best bowler or batsman for a given match situation. The most effective neural network among the several tested was a three hidden layer perceptron. With the use of k-Means and hierarchical clustering, which uncovered exact patterns and limitations, data was created and the best team for a match was predicted [8]. The resulting models were capable of making instantaneous, accurate predictions. The association rule mining techniques led to achieve 68% through the data visualization aspects. This assists to different sets of rules that affects the player's performance [9] eventually. This also pins on the analysis that plays the key role within the team selection part.This paper makes a way that is focused to predict a player's performance through different algorithms like [10] Support Vector Machine and Linear Regression as well.

III. PROBLEM IDENTIFDICATION

Existing many algorithms and approaches to solve one problem. The effectiveness, precision and complexity of these many algorithms and methods may vary. The decision of the methods to use may rely on the elements like the size of the list, the required level of accuracy and the resources available. Each of these algorithms have pros and limitations. No efficient and appropriate usage of the different algorithms with different attributes towards the problem solving. Selecting the appropriate algorithm for the given task is essential for obtaining optimal performance because different algorithms are built with certain properties that make them effective for handling particular sorts of issues. People frequently lack the knowledge or the skills required to choose the algorithm that is most appropriate for a specific reason because they are unfamiliar with all the possible algorithms. This may lead to use of inefficient or faulty algorithm. Confusion and contradiction within the coach and captain to take decisions during a match or before.

TABLE I. PERFORMANCE COMPARISONS OF SEVERAL RESEARCH STUDIES

References	**Data Source**	**Approach**	**Accuracy**	**Evaluation Metrics**
[1]	ESPN cric.info	CNN	70%	F1 Score
[2]	ESPN cric.info	XGBoost	80%	F1 Score, Confusion Matrix
[3]	IPL Website	XG Boost	38%	Error Rate
		Cat Boost	34%	
		Random Forest	45%	
[4]	ESPN cric.info	Random Forest	95.78%	Confusion Matrix
		SVM Classifier	93.64%	
		Decision Tree	91.87%	
[5]	Cricsheet.org	Random Forest	75%	Data Visualization Techniques
[6]	ESPN cric.info	SVM	93%	F1 Score
		Naive Bayes	74%	
		Random Forest	82%	
[7]	ESPN cric.info	Random Forest	Batsmen: 76%	F1 Score
			Bowler:67%	
			All-rounder: 98%	

[8]	Kaggle	K-Means and Hierarchical clustering	56%	Data Visualization Technique
		ANN	46%	
		K-Means	29%	
[9]	Cricbuzz.com	Association Rule Mining	68%	Data Visualization Techniques
[10]	ESPN cric.info	SVM	91.5%	K-fold Cross Validation

Table 1 presented the comparative analysis of several research studies presented in the proposed area of research. It also analyzes the trend in usage of algorithms that are majorly used to make a particular conclusion regarding the performance of the player. The particular relation does not exist between the data source and algorithm used to analyze. However the often used data collection source is ESPN and cric.info websites. As it consists of real time data with true facts as well. Moreover there are several evaluation metrics. Fig.1 depicts the actual workflow of the system which assists the coaches and data analysts to make further decisions and steps.

IV. Proposed Solution

The proposed solution from the analysis made is presented as shown in work flow of the system. It consists of step by step actions that have to be performed during any decision making challenges and predictions. This work flow can assist to have a clear view of the actions to be taken which eventually leads to decision making and predictions as well.

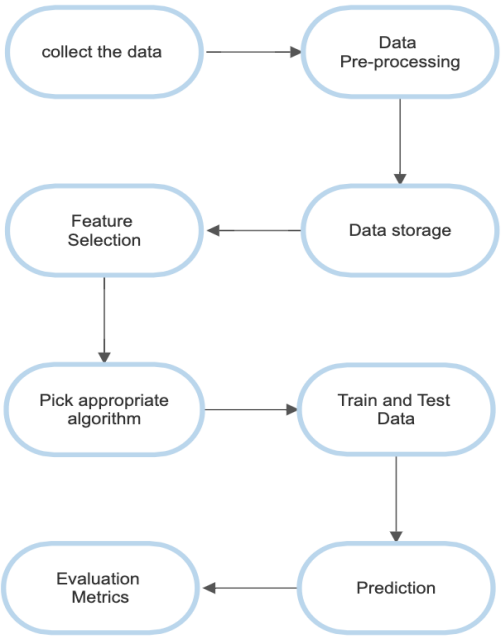

Fig.1 Work-Flow diagram

Fig.2 presents the proposed system architecture for the existing system. It includes the flow of the system and it assists to make squads for upcoming matches as well.

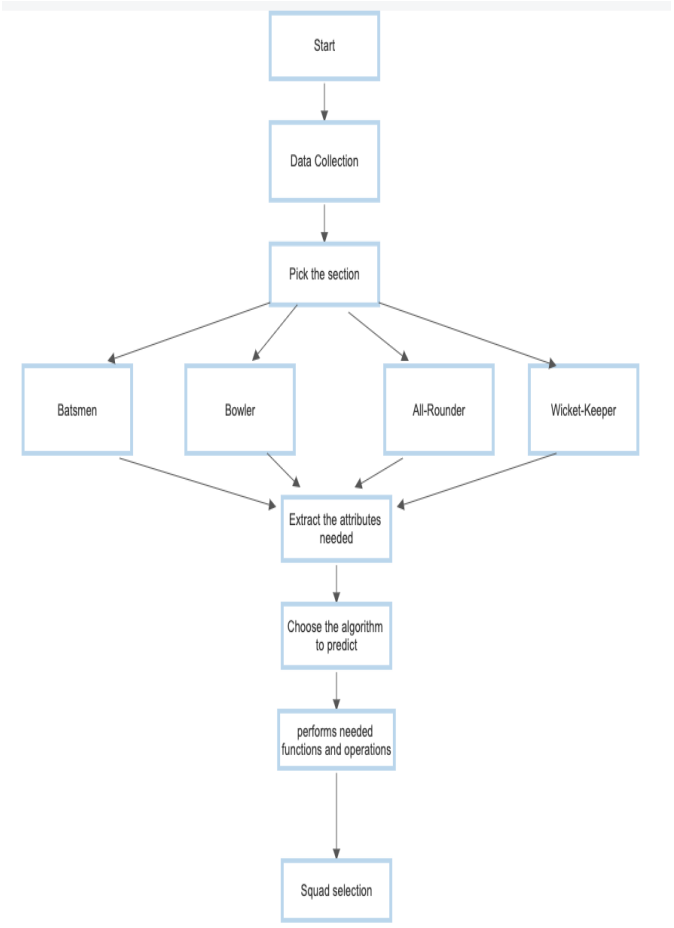

Fig. 2 . Proposed solution for the problem identification

Table. 2 describes the core attributes that has to be involved in the data set while making team selection or the analysis or making different predictions in the different formats of cricket as well. To use the attributes and features mentioned for each type of cricket player in a dataset to make predictions, we can use various machine learning algorithms which mentioned in the Table.1. Nevertheless the process goes as follow as given below.

TABLE II. PERFORMANCE COMPARISONS OF SEVERAL RESEARCH STUDIES

Batsmen	**Bowler**	**All-rounder**	**Wicketkeeper**
Runs scored	Wickets taken	Runs scored	Catches taken
Batting average	Bowling average	Batting average	Stumppings
Strike rate	Economy rate	Strike rate	Dismissals
Innings played	Strike rate	Wickets taken	Matches played
Balls faced	Overs bowled	Bowling average	Innings played
Not outs	Maiden overs	Economy rate	Batting average
Centuries	Five-wickets hauls	Overs bowled	Batting position
Half-centuries	Ten-wickets hauls	Maiden overs	Strike rate
Batting position	Bowling-position	Batting position	Opposition
Opposition	Opposition	Bowling position	Match format
Match format	Match format	Match format	Successful reviews

A. DATA COLLECTION:

Collect relevant data for each cricket player, including attributes such as runs scored, wickets taken, catches taken etc, along with contextual information such as match format, opposition, venue and date likewise attributes which are mentioned in Table 2.

B. DATA PREPROCESSING:

Clean the data by removing any missing values or outliers and normalize or standardize the data to make it easier for the algorithm to process.

C. FEATURE SELECTION:

Choose which features (attributes) to include in the analysis. Not all features may be equally important, so it's important to select only those that have a significant influence on the performance of player.

D. MODEL TRAINING:

Divide the data into training and testing sets, then use the selected algorithm to train a machine learning model on the training set. Based on the chosen features, the model will learn from the data to forecast the player's performance.

E. MODEL EVALUATION:

Test the model on the testing set and compare the predicted values to the actual values to assess the model's performance. Depending on the algorithm being used, the evaluation metrics can change, however some that are frequently employed are the F1 Score, the Confusion Matrix, and others that are listed in Table 1.

F. PREDICTION:

Once the model is trained and evaluated, it can be used to make predictions on new data(i.e., data for a player that was not included in the training or testing sets). The model will use the selected features to predict the player's performance.

By using relevant attributes and features, we can preprocess the data for each type of cricket player and use machine learning algorithms to make predictions about their performance. This can help teams and coaches make data-driven decisions and improve the player's performance.

V. DISCUSSION

Making data-driven decisions for clubs and coaches can be aided by using machine learning to examine cricket player performance. Machine learning algorithms can be trained to forecast a player's performance based on historical data by gathering pertinent data and preparing it with the right traits and features. Many machine learning methods can be used to evaluate cricket players' performance. The particular problem and dataset in question will determine which algorithm is used. Following the analysis, the following algorithms are recommended: neural networks, support vector machines, random forests, and decision trees. F1-Score and confusion matrix are used as evaluation metrics. The best data set source is ESPN cricinfo, cricbuzz which are robust, scalable and it consists of true data. By utilizing pertinent traits and characteristics, we may enhance player selection, create fresh tactics, and raise team productivity.

VI. Conclusion

In conclusion, this comparison study has shown how well machine learning algorithms may be used to forecast and assess cricket player's performance. The findings of this study demonstrate that machine learning techniques can be used to effectively anticipate a player's performance as well as to pinpoint the crucial elements that influence their on-field success. While it offers a useful tool for evaluating player performance and reaching well-informed conclusions about team selection, this research has significant significance for cricket coaches and selectors. The creation of increasingly complex machine learning algorithms as well as the investigation of various data sources and analytical methods may be part of future research in this field. Ultimately, this study has the ability to change how we perceive, understand and analyze it.

References

[1] S. Manivannan and M. Kausik, "Convolutional Neural Network and Feature Encoding for Predicting the Outcome of Cricket Matches," 2019 14th Conference on Industrial and Information Systems (ICIIS), Kandy, Sri Lanka, 2019, pp. 344-349, doi: 10.1109/ICIIS47346.2019.9063316.

[2] A. Ghosh, A. Sinha, P. Mondal, A. Roy and P. Saha, "Indian Premier League Player Selection Model Based on Indian Domestic League Performance," 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), NV, USA, 2021, pp. 0277-0282, doi: 10.1109/CCWC51732.2021.9376011.

[3] S. Choudhari, N. Wagholikar, A. Swaminathan and S. Kurhade, "Dream11 IPL Team Recommendation using Machine Learning and Skill-Based Ranking of Players," 2022 International Conference for Advancement in Technology (ICONAT), Goa, India, 2022, pp. 1-6, doi: 10.1109/ICONAT53423.2022.9725819.

[4] A. Balasundaram, S. Ashokkumar, D. Jayashree and S. Magesh Kumar, "Data mining based Classification of Players in Game of Cricket," 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2020, pp. 271-275, doi: 10.1109/ICOSEC49089.2020.9215413.

[5] V. V. Tharoor and N. M. Dhanya, "Performance of Indian Cricket Team in Test Cricket: A comprehensive Data Science analysis," 2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC), Chennai, India, 2022, pp. 128-133, doi: 10.1109/ICESIC53714.2022.9783492.

[6] M. K. Mahbub, M. A. M. Miah, S. M. S. Islam, S. Sorna, S. Hossain and M. Biswas, "Best Eleven Forecast for Bangladesh Cricket Team with Machine Learning Techniques," 2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Dhaka, Bangladesh, 2021, pp. 1-6, doi: 10.1109/ICEEICT53905.2021.9667862.

[7] M. Shetty, S. Rane, C. Pandita and S. Salvi, "Machine learning-based Selection of Optimal sports Team based on the Players Performance," 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 1267-1272, doi: 10.1109/ICCES48766.2020.9137891.

[8] P. Jhansi Rani, A. Vidyadhar Kamath, A. Menon, P. Dhatwalia, D. Rishabh and A. Kulkarni, "Selection of Players and Team for an Indian Premier League Cricket Match Using Ensembles of Classifiers," 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2020, pp. 1-6, doi: 10.1109/CONECCT50063.2020.9198371.

[9] U. J. Solanki and J. Vala, "Selection for balanced cricket team fourth coming ICC Championship 2017," 2017 2nd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2017, pp. 794-797, doi: 10.1109/CESYS.2017.8321193.

[10] A. I. Anik, S. Yeaser, A. G. M. I. Hossain and A. Chakrabarty, "Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms," 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT), Dhaka, Bangladesh, 2018, pp. 500-505, doi: 10.1109/CEEICT.2018.8628118.

[11] M. Devanandan, V. Rasaratnam, M. K. Anbalagan, N. Asokan, R. Panchendrarajan and J. Tharmaseelan, "Cricket Shot Image Classification Using Random Forest," 2021 3rd International Conference on Advancements in Computing (ICAC), Colombo, Sri Lanka, 2021, pp. 425-430, doi: 10.1109/ICAC54203.2021.9671109.

[12] D. Hasanika, R. Dilhara, D. Liyanage, A. Bandaranayake and S. Deegalla, "Data Mining System for Predicting a Winning Cricket Team," 2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS), Kandy, Sri Lanka, 2021, pp. 92-97, doi: 10.1109/ICIIS53135.2021.9660702.

[13] R. Chaudhary, S. Bhardwaj and S. Lakra, "A DEA Model for Selection of Indian Cricket Team Players," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 224-227, doi: 10.1109/AICAI.2019.8701424.

[14] N. Rodrigues, N. Sequeira, S. Rodrigues and V. Shrivastava, "Cricket Squad Analysis Using Multiple Random Forest Regression," 2019 1st International Conference on Advances in Information Technology (ICAIT), Chikmagalur, India, 2019, pp. 104-108, doi: 10.1109/ICAIT47043.2019.8987367.

[15] C. Deep Prakash, Sanjay Verma,A new in-form and role-based Deep Player Performance,Index for player evaluation in T20 Cricket,Decision Analytics Journal,Volume 2,2022,100025,ISSN 2772-6622,doi10.1016/j.dajour.2022.100025.

[16] S.A. Hydar Abbas, H. Karvannan, V. Prem,Effect of neuromuscular training on functional throwing performance and speed in asymptomatic cricket players,Journal of Bodywork and Movement Therapies,Volume 23, Issue 3,2019,Pages 502-507,ISSN 1360-8592,https://doi.org/10.1016/j.jbmt.2018.02.014