JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Machine Learning Based Prediction of Optimal Batsmen and Bowlers for the Indian Cricket Team 's Playing XI)

Bhaskar S, Nanjundeshwar, V Padmapriya

Data Scientist – I, PG Student, Associate Professor
[1](Data Science Service Line), [2,3]Department of Computer Applications
[1]Mathco,[2,3]B.M.S. College of Engineering Bengaluru, India

Abstract : The Indian national cricket team stands as a symbol of pride and unity for millions of fans. Renowned for its achievements and depth of talent, the team has played a pivotal role in shaping the global cricketing landscape. However, consistently selecting the optimal playing XI across Test, ODI, and T20I formats remains a persistent challenge. This paper presents a data-driven approach to identifying the best combination of batsmen and bowlers for India's playing XI, leveraging recent player performance data and advanced machine learning algorithms. ESPN Cricinfo data sets, such as matches played, batting and bowling averages, strike rate, economy rate, runs, and wickets, form the basis of objective team selection.

A range of supervised machine learning algorithms, like Random Forest, Gradient Boosting, AdaBoost, K-Nearest Neighbors, Support Vector Regressor, and XGBoost, are compared on the basis of standard regression evaluation metrics like Mean Absolute Error, Mean Squared Error, and R^2 Score. The suggested approach provides selectors and coaches with a useful tool to build a balanced and competitive XI, and to aid evidence-based decision-making. This paper underscores the potential of Machine Learning to enhance cricket analytics and offers a scalable approach for modernizing team management in international cricket.

IndexTerms—Data analytics, ensemble models, Indian cricket team, Multi format, playing XI prediction, Sports analytics, Team selection

I. INTRODUCTION

Cricket holds a significant and influential role in Indian society, serving as both a national sport and a unifying force for millions. The Indian national cricket team, admired for its phenomenal success and legendary players, has experienced both moments of triumph and periods of disappointment on the international stage. India's cricketing journey includes historic victories such as the ICC Cricket World Cup wins in 1983 and 2011, the ICC Champions Trophy in 2013, and the inaugural ICC T20 World Cup in 2007 [1]. However, the team has also faced criticism for inconsistent performances, particularly in overseas conditions and during high-pressure tournaments. Notable setbacks include early exits in the 2007 ODI World Cup, the 2016 and 2022 T20 World Cups, and struggles in Test series abroad, especially in countries like South Africa, England, and New Zealand. These fluctuations in success have often been linked to challenges in team selection, over-reliance on a handful of star players, and difficulties in adapting to varied playing conditions. Picking the best set of batsmen and bowlers into the playing XI is a trying and intricate decision that has direct implications for the performance of the team. The process is a careful juggling of player statistics, recent form, fitness, opposition study, and conditions specific to the match. Historically, selectors and coaches used to trust their experience and judgment in making such decisions. While such a level of expertise is worth its weight in gold, it brings a degree of subjectivity and may not always constitute the most up-to-date or applicable data, particularly with the ever-changing state of cricket nowadays.

The field of data analytics and machine learning has paved the way to introduce more objectivity and accuracy to the selection process. The ML algorithms process large amounts of player performance data, determine patterns, and make predictions that can aid in evidence-based decision-making. In Indian cricket, the use of such technologies can aid in overcoming existing challenges in team selection, for example, balancing experienced players with developing talent and responding to varying playing conditions across formats. While these technologies have the potential to make a difference, their use in choosing the playing XI of the Indian cricket team, namely, determining the best combination of batsmen and bowlers, has yet to be fully explored. Research has generally concerned matching game results or assessing the performance of individual players, not consolidating such insights to suggest a proper balance of and competitive team. This paper proposes a Machine Learning-driven approach to predict the best pair of batsmen and bowlers for India's playing XI. Using recent and detailed player performance data, the model seeks to give selectors and coaches data- driven, actionable suggestions. The approach is flexible across all three formats, like the Test, ODI, and T20I, to remain relevant to the changing dynamics of international cricket. The rest of the section is organized as follows: Section II reviews related work in the Literature Survey. Section III describes the Experimentation Setup and Methodology. Section IV presents the Results and Discussion. Finally, Section V presents the Conclusion.

II. LITERATURE SURVEY

Literature review is essential to understand the developments and challenges in cricket analytics, especially in the areas of team selection and performance prediction. Previous research has explored a range of Machine Learning techniques, key performance indicators, and data sources to enhance decision- making in cricket. This section summarizes the most relevant findings and methodologies from earlier studies, providing context and justification for the current work. The use of Machine Learning (ML) and data analytics across cricket has grown exponentially over the past few years due to the increased availability of granular match and player data. Initial research in the field relied on using past performance statistics to forecast player performances and make better team choices. Anand [2], showed that supervised ML models, especially Random Forest, can make highly accurate predictions in classifying and predicting the performances of batsmen, bowlers, and all-rounders in Test cricket, with classification accuracies above 96%. Their paper sheds light on the utility of ensemble methods and the significance of strong data preprocessing and feature engineering. Building on this, Raghunandan, Rao et al. [3] developed a hybrid system based on rule-based algorithms and probabilistic and ML models to make team selection optimal. By using contextual parameters like stadium conditions, pitch type, and opponent analysis, their system makes dynamic suggestions of a balanced playing XI, proving the merit of blending domain knowledge and data-driven information. In a similar manner, Sumathi et al. [4][5] examined the application of linear regression, K-means clustering, and Random Forest for cricket player evaluation and ranking, focusing on the utilization of clustering to discover high- performing clusters and validate the selection of players.

Throughout the literature, the common theme is the discovery of key performance indicators (KPIs) for batsmen and bowlers. Perattur et al. [6][7] laid emphasis on parameters like batting average, strike rate, bowling average, and economy rate, while also recommending the use of sophisticated metrics like boundary percentage, dot ball percentage, and phase-based performance (e.g., powerplay, middle, and death overs). These papers contend that using conventional metrics by themselves is not enough to fully assess players and that context-dependent features greatly improve predictive performance.The IPL has been a fertile ground for ML-powered cricket analytics. Barot, Kothari et.al [8] and Chivate and Manjunath[9] used a variety of algorithms—SVM, Logistic Regression, Random Forest, and XGBoost—to make predictions about match results and analyze player contributions in the IPL. Their results reveal that phase-specific characteristics (e.g., death overs strike rate and middle overs economy) and ensemble models play a significant role in predicting team success.

The studies also point to the practical issues of data quality, selection of features, and the necessity of ongoing model retraining to accommodate changing team lineups and tactics. A number of studies have been directed at match outcome prediction in One Day Internationals (ODIs) and T20s. Haq, Hassan et al. [10],[11] showed that K-Nearest Neighbors, XGBoost, and Linear Regression models can predict match winners and best playing XIs well, with KNN and Random Forest frequently beating other classifiers in precision. Agrawal, Singh et al. [12],[13] also investigated the application of Naïve Bayes, Decision Trees, and XGBoost for predicting IPL matches with accuracies over 90% and highlighting the significance of feature engineering and parameter tuning.

Though the research studies considered here have gone a long way in ascertaining how Machine Learning is improving cricket analytics, they are focused largely on individual tournaments, match formats, or standalone areas of a player's performance. There is still a large gap in the creation of holistic, evidence-based frameworks specific to the distinct needs of the Indian national cricket team, especially in choosing the best selection of batsmen and bowlers across all formats [14]. Current methods tend to neglect the dynamic condition of player form, changing framework of international cricket, and having flexible, evidence-based selection instruments able to react to actual match circumstances.

Despite increasing attention to ML-based team selection, there are various challenges in the Indian context. One of the biggest challenges is combining heterogeneous and contemporary player performance data across different formats—Test, ODI, and T20I—into a single analytical framework [15]. Another issue is the absence of standardized and transparent selection criteria and the use of selectors' subjective judgments, which might result in inconsistency and lost opportunities in team selection. The intricacy of cricket analysis is also augmented by the necessity to cover contextual factors like strength of opposition, pitch state, and recent player form, which are frequently not well-represented by standard models.

Recent experiences with sports analytics and the growing digitization of cricket data indicate that India has a good chance to ride the wave of advanced machine learning methods for more objective and efficient team management. Widespread availability of player detail statistics, the success of data-driven initiatives in home leagues such as the IPL, and the acceptability of analytics by coaches and selectors all indicate that evidence-based decision-making will become the order of the day in Indian cricket in the days to come.

This paper fills these shortcomings by proposing a strong, machine learning-based approach for forecasting the best choice of batsmen and bowlers for the Indian cricket team's playing XI. In contrast to existing research that targets individual tournaments or tournaments of a single format, our solution encompasses recent, multi-format player performance data and uses a set of cutting-edge supervised learning algorithms such as Random Forest, Gradient Boosting, AdaBoost, K-Nearest Neighbors, Support Vector Regressor, and XGBoost. The suggested approach not only analyzes conventional parameters like batting and bowling averages and strike rates but also includes sophisticated, context-sensitive features like phase-based performance and opposition analysis.
In addition, this approach prioritizes transparency and everyday usability through easy-to-interpret rankings and actionable advice for selectors and coaches. By comprehensively analyzing model performance based on traditional regression measures and identifying the most significant features to use in player selection, our study establishes a new standard for Machine Learning in cricket analytics. In the end, this paper aims to render team selection more objective, data-based, and responsive helping propel the Indian national cricket team's continued success on the international scene

III. Experimental Setup and methodology

The experimental setup for this study is grounded in a robust data collection process, which serves as the backbone for further analysis and model development. It is initiated by obtaining extensive player and match datasets for the Indian national cricket team covering all three significant formats: Test, One Day International (ODI), and Twenty20 International (T20I). The main source of this data was ESPN Cricinfo [16], an organizationally well accepted site famous worldwide for providing exhaustive, updated statistics on each international cricket match and player.

A. Data Enhancement

The raw datasets from ESPN Cricinfo included information for all international teams and players. To focus the analysis on the Indian team, the data was curated to include only records pertaining to Indian players and their performances in Test, ODI, and T20I matches over the most recent seasons. The original attributes in the datasets were not always directly suitable for predictive modeling, so key features were derived and selected to best evaluate player performance and team selection.

B. Selected Attributes

The essential attributes chosen for the research reflect both traditional and advanced cricket analytics, and are consistent across all formats to ensure comparability. These include:

Table 1: Batting Attributes

Attribute	Description
Matches Played (Mat)	Total number of matches played by the batsman
Batting Average (Ave)	Number of runs scored per dismissal
Strike Rate (SR)	Number of runs scored per 100 balls faced
Total Runs (Runs)	Aggregate runs scored across matches

Table 1 lists the key statistical features used to evaluate batsmen in the Indian cricket team. These attributes capture both the consistency and scoring ability of each player across different formats.

Table 2: Bowling Attributes

Attribute	Description
Matches Played (Mat)	Total number of matches played by the bowler
Bowling Average (Ave)	Number of runs conceded per wicket taken
Economy Rate (Econ)	Number of runs conceded per over bowled
Total Wickets (Wkts)	Aggregate wickets taken across matches

Table 2 presents the primary metrics considered for assessing bowlers' performance. The selected attributes reflect both wicket-taking capability and efficiency in run containment.

C. Attributes List

The final list of attributes used in the analysis includes:

- For batting: Mat, Ave, SR, Runs
- For bowling: Mat, Ave, Econ, Wkts

Key Attributes:
Mat = Matches, Ave = Average, SR = Strike Rate, Econ = Economy Rate, Runs = Total Runs, Wkts = Total Wickets

D. Proposed Methodology

The methodology employed to predict the optimal playing XI for the Indian national cricket team is designed as a systematic, multi-stage pipeline.

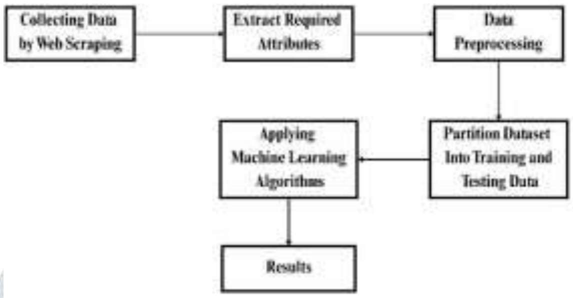


Figure 1: Proposed Methodology.

Figure 1 illustrates the overall workflow of the proposed methodology, starting from data collection and attribute extraction to preprocessing, model training, and result generation. This systematic pipeline ensures a structured approach for predicting the optimal playing XI using machine learning techniques.

Next Sub-Sections from E to I explains about the Methodology employed to predict the top Batsman's and Bowlers for the Indian Cricket Team

E. Collecting Data by Web Scraping

The initial step in the proposed methodology involves gathering relevant cricket data through automated means. Cricket statistics were collected systematically exclusively from ESPN Cricinfo, the official and main database employed in this study. Web scraping scripts implemented through the Python library BeautifulSoup help efficient extraction of an extensive list of match and player statistics [17]. Web scraping is performed by incorporating historical data so that the dataset reflects recent player form, trends, and performance for all three formats, i.e., Test, ODI, and T20I. Quantitative attributes like runs scored, wickets, batting and bowling average, strike rate, economy rate, and matches played, and textual data like the name of the player, role played in the match, type of match, opponent team, and grounds are retrieved.

F. Extract Required Attributes

Having prepared the data, the second step is to choose and filter out the most relevant features to feed into the ML models. The optimal features most opt for player evaluation and team formation are derived and engineered. The data is preprocessed by eliminating redundant columns like player names, match IDs, and redundant metadata, and feature naming and formatting are universalized for all data types.

G. Data Pre Processing

Data preprocessing is carried out on the extracted dataset to improve its quality and make it ready for modeling. All the numeric columns are converted into suitable data types, and missing and incomplete data points are dealt with mean imputation. Only records related to Indian players are preserved, and any duplicate or inconsistent records are discarded to ensure data integrity.

H. Partition Dataset into Training and Testing Data

To enable robust model evaluation and prevent overfitting, the preprocessed dataset is divided into training and testing subsets. An 80:20 split is typically employed, with 80 percent of the data allocated for model training and 20 percent reserved for testing. Stratified sampling is utilized when necessary to ensure that the distribution of key features, such as player positions or match types, remains consistent across both subsets.

I. Applying Machine Learning Algorithms

After partitioning the data and cleaning, various supervised machine learning algorithms were utilized to predict player performance and suggest the optimal playing XI. The models utilized are Random Forest, Gradient Boosting, AdaBoost, K-Nearest Neighbors, Support Vector Regressor, and XGBoost. Each model was tested with common regression scores like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 Score to provide a strong comparison for batting as well as bowling predictions.
A brief overview of the machine learning algorithms employed in this paper is provided below to highlight their key characteristics and suitability for the prediction tasks. Random Forest relies on a committee of decision trees that are overfitting-resistant and

suitable for complex data. Gradient Boosting develops models in sequence to fix past mistakes and is good at identifying detailed patterns. AdaBoost enhances accuracy by concentrating on difficult- to-predict instances through weighted resampling. K-Nearest Neighbors makes predictions based on the mean of the nearest neighbors in feature space [21]. Support Vector Regressor uses kernel functions to capture both linear and non-linear relationships. XGBoost is a high-performance version of gradient boosting that is renowned for its speed, predictive capability, and regularization to avoid overfitting [18][19][20][21][22][23].

IV. RESULTS AND DISCUSSIONS

This section describes the results of different machine learning algorithms used for predicting the players in the Indian national cricket team. The respective analysis was conducted for both batting and bowling using the Test cricket dataset. Multiple regression models were evaluated based on Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 Score

A. Model Performance Comparison

Table 3 presents the comparative performance of all models for both batting and bowling prediction tasks.

Table 3: Model Performance Metrics for Batting and Bowling

Model	Batting MAE	Batting MSE	Batting R^2	Bowling MAE	Bowling MSE	Bowling R^2
Random Forest	74.14	25097.76	0.59	4.82	54.79	0.71
Support Vector Regressor	165.06	79115.86	-0.29	9.48	216.91	-0.13
Gradient Boosting	69.30	21374.50	0.65	4.23	43.93	0.77
K-Nearest Neighbors	88.93	29837.25	0.51	4.11	42.82	0.78
AdaBoost	79.33	26527.10	0.57	5.15	55.55	0.71
XGBoost	**64.01**	**18862.59**	**0.69**	**3.17**	**34.60**	**0.82**

Of all the models that were tested, XGBoost achieved the highest R^2 value for both batting (0.69) and bowling (0.82) performance prediction, indicating superior predictive accuracy. Gradient Boosting and K-Nearest Neighbors also performed well, while Random Forest and AdaBoost produced competitive results. In contrast, the Support Vector Regressor performed very poorly, with negative R^2 values in both tasks. Table 3 Shows the impact of Regression Algorithms especially XGBoost Algorithm in Prediction Tasks

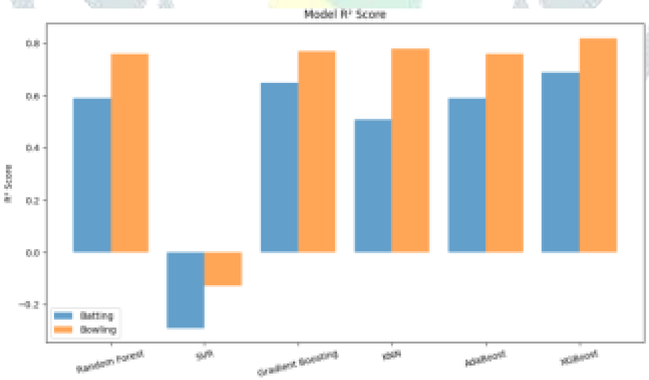


Figure 3: R^2 Comparison of Batting and Bowling Performance Across all Models

Figure 3 presents the R^2 scores achieved by various machine learning models for both batting and bowling performance prediction. The results clearly show that ensemble models such as Random Forest, Gradient Boosting, AdaBoost, and XGBoost consistently yield higher R^2 values for both tasks, reflecting their superior predictive accuracy. In contrast, the Support Vector Regressor (SVR) demonstrates poor performance, with negative R^2 scores for both batting and bowling. Among all models, XGBoost achieves the highest R^2 scores, making it the most effective model for forecasting both batting and bowling performance in this study. These findings underscore the importance of using advanced ensemble techniques for reliable player selection in the Indian cricket team. A higher R^2 value indicates better model predictions and, consequently, more accurate identification of the best players for the Indian Cricket Team.

B. Feature Importance Analysis

Feature importance analysis using the XGBoost model was conducted to identify the most influential factors contributing to player performance prediction.

Table 4: XGBoost Feature Importances for Batting Performance

Feature	Importance
Matches played (Mat)	0.5499
Batting average (Ave)	0.4103
Strike rate (SR)	0.0397

Table 5: XGBoost Feature Importances for Bowling Performance

Feature	Importance
Matches played (Mat)	0.6879
Bowling average (Ave)	0.3017
Economy (Econ)	0.0104

Tables 4 and 5 presents the feature importances derived from the XGBoost model for batting and bowling performance, respectively. In both cases, matches played and average are the most influential predictors for the Indian cricket team, while strike rate and economy rate have comparatively lower impact. These results indicate that experience (matches played) and average (batting or bowling) are the most influential predictors of player performance in the Indian cricket team, while strike rate and economy rate also contribute to the overall predictions.

A. XGBoost Hyperparameter Tuning

Given its superior performance in the initial analysis, XGBoost was further optimized through hyperparameter tuning [24] using GridSearchCV and evaluated with cross- validation and a held-out test set.

Table 6: Tuned XGBoost Hyperparameters and Cross-Validation R^2 Scores

Task	Learning rate	Max depth	n estimators	CV R^2 Score
Batting	0.1	3	100	0.8141
Bowling	0.1	5	100	0.7745

Table 6 shows that hyperparameter tuning enabled XGBoost to achieve strong cross-validated R^2 scores for both batting (0.8141) and bowling (0.7745), confirming its effectiveness for player performance prediction in the Indian cricket team.

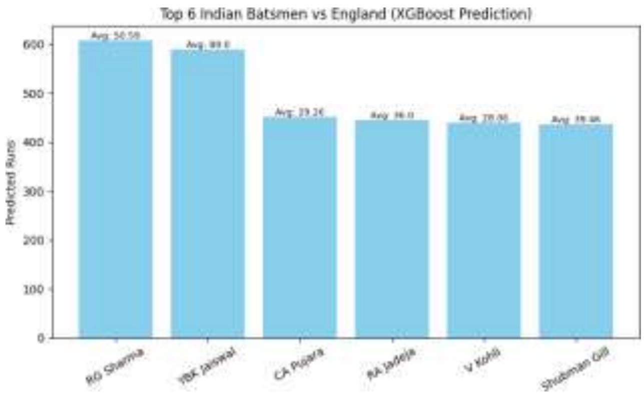
The results demonstrate that XGBoost, especially after hyperparameter tuning, provides robust and accurate predictions for both batting and bowling performance. Its strong cross-validated R^2 scores highlight its reliability for player evaluation. Overall, XGBoost stands out as an effective tool for data-driven decision-making in the Indian cricket team.

The same process and feature set were applied to the ODI and T20 formats, and comparable patterns in model performance were observed. XGBoost consistently outperformed the other models across all formats, confirming the robustness and generalizability of the proposed methodology.

A. Validation Against Actual Team Selections

To determine the practicality of the suggested approach, the highest Indian batsmen and bowlers forecasted by the XGBoost model for the England series were contrasted with the announced Indian Test team for 2023-24 [25].

The announced team consisted of: Rohit Sharma (C), Shubman Gill, Yashasvi Jaiswal, Virat Kohli, Shreyas Iyer, KL Rahul (WK), KS Bharat (WK), Dhruv Jurel (WK), Ravichandran Ashwin, Ravindra Jadeja, Axar Patel, Kuldeep Yadav, Mohd. Siraj, Mukesh Kumar, Jasprit Bumrah (VC), and Avesh Khan.



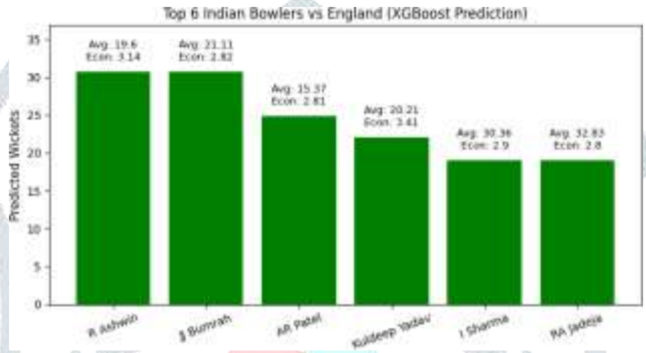


Figure 4: XGBoost top 6 Indian batsmen against England.

Figure 5: XGBoost top 6 Indian bowlets against England.

Figures 4 and 5 present the XGBoost-predicted top 6 Indian batsmen and bowlers against England. Predicted runs and batting averages are shown for each batsman, while predicted wickets, bowling averages, and economy rates are displayed for each bowler

Table 6: Comparison of XGBoost Model-Predicted Top Performers and Actual Indian Test Squad

Player	Model Predicted	In Actual Squad
RG Sharma	Yes	Yes
YBK Jaiswal	Yes	Yes
CA Pujara	Yes	**No**
RA Jadeja	Yes	Yes
V Kohli	Yes	Yes
Shubman Gill	Yes	Yes
R Ashwin	Yes	Yes
JJ Bumrah	Yes	Yes
AR Patel	Yes	Yes
Kuldeep Yadav	Yes	Yes
I Sharma	Yes	**No**

Table 6 highlights the strong alignment between the model's predictions and the actual Indian Test squad, demonstrating the effectiveness of the proposed machine learning approach in identifying key players for team selection.

The model's predictions closely match the actual Indian Test squad, accurately identifying key players. This strong alignment demonstrates the effectiveness of the machine learning approach in supporting objective and data-driven team selection for Indian cricket.

V. RESULTS AND DISCUSSIONS

The data-driven prediction framework for the choice of the best batsmen and bowlers to be included in the Indian cricket team's playing XI, as discussed in this paper, is a major milestone in the development of data-driven team choice. Unlike older methods, which tend to be based on personal opinion or narrow performance indicators, this paper uses a wide-ranging set of player statistics, context-rich match data, and sophisticated ensemble techniques to provide objective, actionable team composition insights. By leveraging diverse player statistics and advanced ensemble machine learning techniques, the proposed framework offers objective, actionable insights for selecting the optimal playing XI, moving beyond traditional subjective methods. The results demonstrate strong alignment with actual team selections, underscoring the practical value of machine learning in modern team management.

However, it is important to note that the current methodology does not account for certain real-world constraints such as player injuries, workload management, or team balance (e.g., left/right-hand pairs, all-rounders, wicketkeepers). Contextual factors like pitch conditions, weather, and opposition analysis are also not included due to data limitations, and the model does not dynamically update for last-minute changes or real-time events. Addressing these limitations will be a key focus for future research, with the aim of further enhancing the robustness and real-world applicability of the framework. In Conclusion, this paper not only illustrates the potential of machine learning to optimize Indian cricket team selection but also raises a new standard for transparency, flexibility in cricket analytics. By embracing such cutting-edge technologies, Indian cricket will be able to continue to excel in its tradition and retain a competitive advantage on the global platform.

VI. REFERENCES

I. "India's World Cup wins: 1983, 2011 and the rise of a cricketing superpower," ESPN Cricinfo, [Online]. Available: https://www.espncricinfo.com/story/india-s-world-cup-wins-1983- 2011-and-the-rise-of-a-cricketing-superpower-1252267

II. Anand, Sourav, Nitish Kumar, and Vaneet Kumar. "Machine Learning- Based Performance Prediction Models for Test Cricketers: A Data- Driven Approach." In 2023 2nd International Conference on Futuristic Technologies (INCOFT), pp. 1-6. IEEE, 2023.

III. NS, Krishnaraj Rao, K. R. Raghunandan, Shamanth Rao, Samarth Patil, Sai Nandan, and Rathay Devadiga. "Cricket Player's Performance Prediction Using Machine Learning." In 2025 International Conference on Artificial Intelligence and Data Engineering (AIDE), pp. 944-950. IEEE, 2025.

IV. Sumathi, M., S. Prabu, and M. Rajkamal. "Cricket players performance prediction and evaluation using machine learning algorithms." In 2023 international conference on networking and communications (ICNWC), pp. 1-6. IEEE, 2023.

V. Perattur, Nagabushanam, K. Vasanth, Vivek Singh Kushwah, Thota Praneeth Sai, S. Rekha, and Radha Subramanyam. "Algorithms for Performance Prediction in Cricket–Factors Influencing the Player." In 2024 IEEE 16th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 1047-1051. IEEE, 2024.

VI. Prakash, C. Deep, and Sanjay Verma. "A new in-form and role-based deep player performance index for player evaluation in T20 cricket." Decision Analytics Journal 2 (2022): 100025.

VII. Barot, Harshit, Arya Kothari, Pramod Bide, Bhavya Ahir, and Romit Kankaria. "Analysis and prediction for the Indian Premier League." In 2020 International Conference for Emerging Technology (INCET), pp. 1-7. IEEE, 2020.

VIII. Chivate, Chirag Mahaveer, and D. R. Manjunath. "Data-Driven Strategies for Enhancing Performance in Indian Premier League: A Case Study of Royal Challengers Bangalore." In 2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS), pp. 1-6. IEEE, 2024.

IX. Haq, Inam Ul, Inzimam Ul Hassan, and Hilal Ahmad Shah. "Machine learning techniques for result prediction of One Day International (ODI) cricket match." In 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), pp. 1-5. IEEE, 2023.

X. Ravichandran, Nilesh, Nihal Chengappa P.A, and Pranav Rao Rebala. "Optimal IPL Playing 11 Team Selection." In 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), pp. 1-6. IEEE, 2023.

XI. Agrawal, Shilpi, Suraj Pal Singh, and Jayash Kumar Sharma. "Predicting results of Indian premier league T-20 matches using machine learning." In 2018 8th International Conference on Communication Systems and Network Technologies (CSNT), pp. 67-71. IEEE, 2018.

XII. Vistro, Daniel Mago, Faizan Rasheed, and Leo Gertrude David. "The cricket winner prediction with application of machine learning and data analytics." International journal of scientific & technology Research 8, no. 09 (2019): 21-22.

XIII. CA, Anser Pasha, B. J. Shishira, Sumedh M. Anvekar, and Uday Rao CG. "Optimal Cricket Team Selection Using Machine Learning." In 2024 Second International Conference on Advances in Information Technology (ICAIT), vol. 1, pp. 1-5. IEEE, 2024

XIV. Patel, S., Panchal, V. & Shah, D. (2022). "A unified Bayesian hierarchical model for multi-format player valuation in international cricket." Journal of Sports Analytics 8(3), 211-228. DOI:10.3233/JSA- 220537

XV. Vistocco, M., Saikia, H. & Scarf, P. (2021). "Combining performance metrics across formats: A weighted-plus-minus approach applied to Indian players." Proceedings of the MathSport International Conference, 45-52. ISBN:978-88-6938-112-4

XVI. ESPN Cricinfo, "Statsguru: Cricket statistics database," [Online]. Available: https://www.espncricinfo.com/

XVII. L.Richardson, "Beautiful Soup Documentation," [Online]. Available: https://www.crummy.com/software/BeautifulSoup/

XVIII. L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

XIX. J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics, vol. 29, no. 5, pp. 1189-1232, 2001.

XX. Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119-139, 1997.

XXI. T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21-27, 1967.

XXII. A. J. Smola and B. Schölkopf, "A Tutorial on Support Vector Regression," Statistics and Computing, vol. 14, no. 3, pp. 199-222, 2004.

XXIII. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016.

XXIV. https://scikit-learn.org/stable/modules/grid_search.html

XXV. https://www.bcci.tv/articles/2024/news/55556084/india-s-squad-for- first-two-tests-against-england-announced