IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 12, Issue 3, May-June 2025

Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203013

Cricket Team Prediction using Machine Learning

E. Vikash, Dr. K. Poornapriya, Dr. G. Aarthi, B. Kowsalya

Department of Master of Computer Applications, Vidyaa Vikas College of Engineering and Technology, Tiruchengode, Tamil Nadu, India

Principal, Vidyaa Vikas College of Engineering and Technology, Tiruchengode, Tamil Nadu, India

Head of the department, Department of Master of Computer Applications, Vidyaa Vikas College of Engineering and Technology, Tiruchengode, Tamil Nadu, India

Assistant Professor, Department of Master of Computer Applications, Vidyaa Vikas College of Engineering and Technology, Tiruchengode, Tamil Nadu, India

ABSTRACT: The cricket team selection process presented in this study combines batting and bowling datasets, leveraging various performance metrics to identify the most suitable players. Each cricketer is evaluated using multiple parameters and categorized into performance levels such as Excellent, Very Good, Good, Satisfactory, and Poor. Prominent batsmen like Shikhar Dhawan, Virat Kohli, Rohit Sharma, and MS Dhoni are ranked as Excellent, whereas players like Ishan Kishan and Sanju Samson fall under the Poor category. In the bowling dataset, standout performers include Ravindra Jadeja, Bhuvneshwar Kumar, and Jasprit Bumrah. Feature selection algorithms like Particle Swarm Optimization (PSO) and Correlation-based Feature Selection (CS) are utilized to identify critical attributes such as batting average, runs scored, bowling average, and opponent-specific performance. These refined features enhance the accuracy of evaluating player capabilities. Consequently, top-performing players like Virat Kohli and Ravindra Jadeja are selected for their consistent and impactful contributions to their respective teams.

KEYWORDS: Cricket, Performance Prediction, Machine Learning, Decision Tree, Random Forest

I. INTRODUCTION

Intelligent cricket team selection plays a crucial role in achieving success in the highly competitive world of cricket. Building a team that strategically balances individual player strengths and weaknesses can greatly influence match outcomes. The integration of machine learning techniques offers a powerful way to predict player performance with notable accuracy. This approach involves key processes such as data acquisition, preprocessing, feature engineering, and model selection. By utilizing historical data, including player statistics, match conditions, and outcomes, predictive models are developed to forecast future performances. These models help selectors and coaches make informed decisions by considering factors like recent form, pitch conditions, and opponent-specific performance. Ultimately, this data-driven approach enables the creation of a well-rounded team with the best chance of securing victory.

1.1 CRICKET

Cricket is a sport steeped in tradition and renowned for its rich history, captivating matches, and passionate following in many parts of the world. It is a bat-and-ball game that is played between two teams, each comprising 11 players. Cricket is not just a sport but a cultural phenomenon that has captured the hearts of millions. The game's origins can be traced back to the 16th century in England, but it has since evolved into a global sport with diverse formats, including Test cricket, One Day Internationals (ODIs), and Twenty20 (T20) matches. Each format offers a unique spectacle, ranging from the strategic and endurance-based Test matches, where matches can last up to five days, to the fast-paced and explosive T20 matches that can conclude in a few hours. A cricket match typically involves two phases: batting and bowling/fielding. The team that bats aims to score runs by hitting the ball delivered by the bowler and running between two sets of wooden stumps. The fielding team, on the other hand, strives to dismiss the batsmen by getting them out through various means, such as catching the ball, bowling the batsman out, or running them out.

1.2 PERFORMANCE PREDICTION

Performance prediction, in the realm of data analysis and machine learning, is a crucial and dynamic field with a wide range of applications across various industries. It involves using historical data and statistical models to forecast future outcomes, behaviours, or trends. This predictive capability enables organizations and individuals to make informed

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203013

decisions, optimize resource allocation, and anticipate opportunities and challenges. Performance prediction finds its utility in diverse domains, including finance, healthcare, marketing, sports, and manufacturing. It allows businesses to anticipate customer preferences, financial market movements, equipment failures, and much more. Moreover, it plays a pivotal role in personalization, risk management, and resource optimization. In this age of big data and advanced analytics, performance prediction leverages sophisticated machine learning algorithms, neural networks, and statistical techniques to extract patterns and insights from vast datasets.

1.3 MACHINE LEARNING

Machine Learning, often abbreviated as ML, is a transformative field of artificial intelligence (AI) that has revolutionized the way computers learn and make decisions. It empowers machines to analyse vast datasets, recognize patterns, and make predictions or decisions without being explicitly programmed. Machine learning has rapidly gained prominence across numerous industries, shaping the future of technology and data-driven decision-making. At its core, machine learning simulates human learning processes, enabling computers to improve their performance on specific tasks through experience. These tasks span a wide spectrum, from image recognition and natural language processing to autonomous driving and medical diagnosis. The applications of machine learning are virtually limitless, ranging from improving customer experiences and optimizing supply chains to enhancing healthcare and making breakthroughs in scientific research. Machine learning continues to evolve, pushing the boundaries of what's possible and driving innovation across industries. As it matures, its impact on society, business, and technology is set to grow exponentially, making it an integral part of our increasingly data-centric world. In the following sections, we delve into the fundamentals, methodologies, and real-world applications of machine learning.

1.4 DECISION TREE

Decision trees are a fundamental and powerful tool in the field of machine learning and data analysis. They provide a visual and intuitive way to make decisions or predictions based on input data by recursively partitioning the data into subsets and assigning outcomes or values to each subset. Decision trees are widely used in various domains, including business, healthcare, finance, and engineering, for tasks such as classification and regression. The concept of a decision tree can be likened to a flowchart, where each internal node represents a decision or a test on an attribute, each branch represents an outcome of that test, and each leaf node represents a final decision or prediction. Decision trees are constructed through a process of recursive partitioning, where the dataset is split into subsets based on the values of the input features, and decisions are made at each split point.

1.4 RANDOM FOREST

Random Forest is a versatile and powerful machine-learning algorithm that has gained widespread popularity across various domains for its ability to provide accurate and robust predictions. It is an ensemble learning technique, meaning it combines the predictions of multiple individual models (decision trees) to make more reliable and accurate predictions. Random Forest is particularly renowned for its effectiveness in classification and regression tasks and is used in applications ranging from finance to healthcare to image recognition. At its core, a Random Forest is a collection of decision trees, each of which is trained independently on a subset of the data and a subset of the features. These individual decision trees, often referred to as "weak learners," are relatively simple and prone to over fitting the training data. However, when combined into a "forest" and their predictions are aggregated the ensemble becomes a "strong learner" that generalizes well to new, unseen data.

II. LITERATURE REVIEW

2.1 INCREASED PREDICTION ACCURACY IN THE GAME OF CRICKET USING MACHINE LEARNING

Kalpdrum Passimet.al. Has proposed in this paper, Player selection is one the most important tasks for any sport and cricket is no exception. The performance of the players depends on various factors such as the opposition team, the venue, his current form etc. The team management, the coach and the captain select 11 players for each match from a squad of 15 to 20 players. They analyze different characteristics and the statistics of the players to select the best playing 11 for each match. Each batsman contributes by scoring maximum runs possible and each bowler contributes by taking maximum wickets and conceding minimum runs. This paper attempts to predict the performance of players as how many runs will each batsman score and how many wickets will each bowler take for both the teams. Both the problems are targeted as classification problems where number of runs and number of wickets are classified in different ranges. We used naïve Bayes, random forest, multiclass SVM and decision tree classifiers to generate the prediction models for both the problems. Random Forest classifier was found to be the most accurate for both the problems.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203013

2.2 ANALYSIS OF PERFORMANCE OF BOWLERS USING COMBINED BOWLING RATE

Dibyojyoti Bhattacharjeeet.al. has proposed in this paper Bowling and batting have always been the most important aspects of the game of cricket. But with advent of the newest version of cricket with twenty overs-a-side the task of the bowlers are believed to have become difficult. This paper attempts to use the combined bowling rate to quantify the performance of a sample of bowlers who participated in the fourth season of the Indian Premier League. Based on relevant literature a set of predictors that can influence the performance of the bowlers were determined. Multiple linear regression technique was then used to identify the reasons that are empirically responsible for the performance of the bowlers. It was also observed that experience of the bowler and combined bowling rate at Tewnty20 internationals are the most significant predictors that are responsible for the performance of bowlers in IPL.In the game of cricket, the batsman shall try to score as many runs as he can, without losing his wicket and the bowler's job is to get the batsman out i.e. to dismiss him. In case of limited over games, in addition to dismissing a batsman, the bowler may also try to restrict the batsman from scoring, because the bowler has to bear with a tradeoff, that he should not give away many runs while experimenting in getting the batsmen out.

2.3 PREDICTION OF ATHLETES PERFORMANCE USING NEURAL NETWORKS: AN APPLICATION IN CRICKET TEAM SELECTION

Subramanian Rama Iyer et.al. Has proposed in this system, Team selection for international sports competitions requires predicting performance of individual athletes. We explore the use of neural networks to rate players and select specific players for a competition. We take cricket as an example. Cricket is a game with mass following in British Commonwealth Countries as well as some other countries. National teams visit other countries for bilateral matches as well as play in World Cup tournaments. We employ neural networks to predict each cricketer's performance in the future based upon their past performance. We classify cricketers into three categories – performer, moderate and failure. We collected data on cumulative player performance from 1985 onwards until the 2006–2007 season. The neural network models were progressively trained and tested using four sets of data. The trained neural network models were then applied to generate a forecast of the cricketer's near term performance. Based on the ratings generated and by applying heuristic rules we recommend cricketers to be included in the World Cup 2007.

2.4 PREDICTION OF RISING STARS IN THE GAME OF CRICKET

Haseeb Ahmadet.al. Has proposed in this system Online social databases are rich sources to retrieve appropriate information that is subsequently analyzed for forthcoming trends prediction. In this paper, we identify rising stars in cricket domain by employing machine learning techniques. More precisely, we predict rising stars from batting as well as from bowling realms. For this intent, the concepts of co-players, team, and opposite teams are incorporated and distinct features along with their mathematical formulations are presented. For classification purpose, generative and discriminative machine learning algorithms are employed, and two models from each category are evaluated. As a proof of applicability, the proposed approach is validated experimentally while analyzing the impact of individual features. Besides, model and category wise assessment is also performed. Employing cross validation, we demonstrate high accuracy for rising star prediction that is both robust and statistically significant. Finally, ranking lists of top ten rising cricketers based on weighted average, performance evolution, and rising star scores are compared with the international cricket council rankings.

2.5 PLAYER'S PERFORMANCE PREDICTION IN ODI CRICKET USING MACHINE LEARNING ALGORITHMS

Almetwally M. Mostafatoet.al. Has proposed in this system, This paper presents a method that is aimed towards predicting a cricket player's upcoming match performance by implementing machine learning algorithms. The proposed model consists of statistical data of players of Bangladesh national cricket team which has been collected from trusted sports websites, feature selection algorithms such as recursive feature elimination and univariate selection and machine learning algorithms such as linear regression, support vector machine with linear and polynomial kernel. To implement the proposed model, the accumulated statistical data is processed into numerical value in order to implement those in the algorithms. Furthermore, aforementioned feature selection algorithms are applied for extracting the attributes that are more related to the output feature. Additionally, the machine learning algorithms are used to predict runs scored by a batsman and runs considered by a bowler in the upcoming match. The experimental setup demonstrates that the model gives up to 91.5% accuracy for batsman Tami and up to 75.3% accuracy for bowler Mahmudullah whereas prediction accuracy for other players are also up to the mark.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203013

III. EXISTING SYSTEM

In any game, selection of best players in a team plays vital role in overall team performance. The team selection in any sport is the key task to ensure good performance of the team. Players are selected based on different criteria. In game of cricket selection of players should consider parameters like players own performance, ground condition, weather forecasting, opposition strength and weakness etc. Machine learning can play vital role in players' performance prediction. Machine learning uses historical data of team performance and past performance of individual players to predict overall performance of team. Prediction of individual player performance helps in team building process. Recently many researchers proposed model for prediction of player's performance for a game of cricket. Researchers' uses machine learning approach for prediction. However existing studies omits some vital features related to ground and weather in their study, which have potential to make huge impact on player's performance. We performed detailed study and literature survey to propose efficient performance prediction of players for game of cricket. Our model will help in best team selection and thus improves overall team performance.

IV. PROPOSED SYSTEM

The proposed system integrates a comprehensive approach to cricket team selection, leveraging advanced techniques in data collection, pre-processing, feature selection, and player selection. Through meticulous data collection and pre-processing, we ensure the quality and reliability of batting and bowling statistics, addressing issues like missing values and outliers. Feature selection algorithms like Particle Swarm Optimization (PSO) and Correlation-based Feature Selection (CS) are then employed to identify the most relevant performance metrics for batting and bowling. This enables us to evaluate players based on key factors such as average runs scored, bowling average, and strike rate. Subsequently, top-performing batsmen and bowlers are selected for the team based on their rankings derived from the identified key features. Finally, a sophisticated player selection algorithm considers both individual performances and team constraints to assemble a well-balanced team comprising the optimal combination of batsmen, bowlers, and all-rounders. This holistic approach ensures the selection of a competitive cricket team capable of achieving success across various match scenarios.

4.1 DATA COLLECTION AND PRE-PROCESSING

In this module, we first gather comprehensive batting and bowling statistics of cricket players from reliable sources. These statistics encompass various performance metrics such as runs scored, batting average, bowling average, strike rate, and more. Once the data is collected, we pre-process it to ensure its quality and reliability. This involves handling missing values, detecting and managing outliers, and normalizing the data if necessary to ensure uniformity and comparability across different metrics and players.

4.2 FEATURE SELECTION

Feature selection plays a crucial role in identifying the most relevant and influential metrics for both batting and bowling performances. Utilizing sophisticated algorithms like Particle Swarm Optimization (PSO) and Correlation-based Feature Selection (CS), we sift through the vast array of metrics to pinpoint those that have the most significant impact on player performance. For batting, selected features may include average runs scored, highest score, number of not outs, etc., while for bowling, metrics like bowling average, strike rate, maiden overs, etc., are considered.

4.3 BATSMEN SELECTION

Once the key features for batting are identified, we apply them to evaluate the performance of batsmen. Using the selected metrics, we rank batsmen based on their performance, considering factors like consistency, scoring ability, and contribution to the team's success. The top-performing batsmen are then selected for the team, ensuring a balanced and formidable batting lineup capable of achieving desired outcomes.

4.4 BOWLERS SELECTION

Similar to batsmen selection, the module for bowlers involves applying the selected features to evaluate their performance. Bowlers are assessed based on metrics such as bowling average, strike rate, economy rate, and their ability to take crucial wickets. The identified key features help rank bowlers according to their effectiveness, and the top-performing bowlers are chosen for the team, ensuring a potent bowling attack capable of restricting opponents and taking wickets consistently.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203013

4.5 PLAYER SELECTION ALGORITHM

In the final module, a sophisticated player selection algorithm is employed to assemble the best combination of players for the team. This algorithm, which may utilize techniques like PSO or CS, considers not only the individual performances of batsmen and bowlers but also team constraints such as the required number of batsmen, bowlers, and all-rounders. By assigning weighted scores to each player based on their performance and fulfilling team requirements, the algorithm facilitates the selection of a well-balanced and competitive team capable of excelling in various match scenarios.

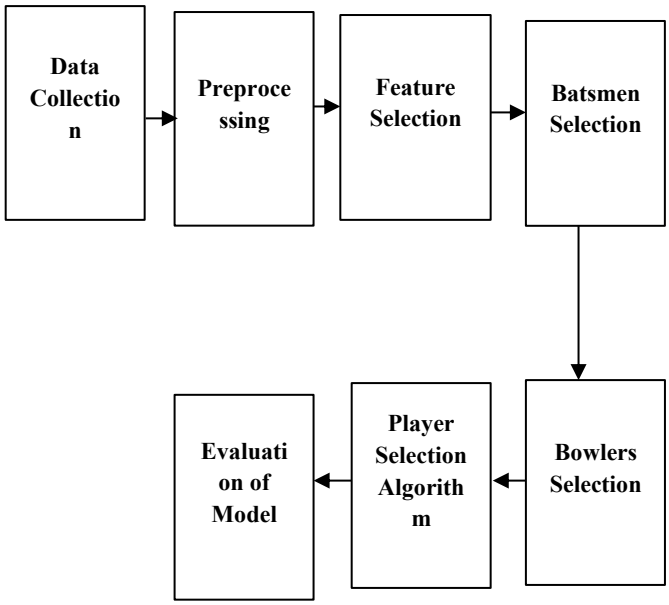

FIGURE 1. SYSTEM FLOW DIAGRAM

V. RESULT ANALYSIS

The result analysis of the batting and bowling datasets reveals significant variations in player performances, with some athletes consistently achieving excellent scores while others fall short. Noteworthy batsmen such as Virat Kohli, Rohit Sharma, and Shikhar Dhawan have shown outstanding consistency, whereas players like Ishan Kishan and Sanju Samson have struggled to maintain form. In the bowling department, Ravindra Jadeja, Bhuvneshwar Kumar, and Jasprit Bumrah have delivered exceptional performances, proving crucial to their teams' success. Conversely, some bowlers have only managed satisfactory or poor performances, highlighting the need for improvement. Feature selection using Particle Swarm Optimization (PSO) and Correlation-based Selection (CS) has helped pinpoint key attributes influencing performance. For batsmen, factors like the number of not outs, highest score, and average were important, while for bowlers, maiden overs, bowling average, and venue-based performance played a key role. These insights have significantly enhanced prediction accuracy in identifying top-tier players. As a result, the final selection includes consistent performers like Virat Kohli, MS Dhoni, Ravindra Jadeja, and Bhuvneshwar Kumar. Overall, the findings emphasize the effectiveness of the selection model and offer valuable guidance for future team-building strategies.

VI. CONCLUSION

In conclusion, the proposed system for cricket team selection embodies a holistic and data-driven approach, integrating advanced techniques in data preprocessing, feature selection, and player selection. By leveraging comprehensive batting and bowling statistics and employing sophisticated algorithms, we ensure the identification of top-performing players based on key performance metrics. The systematic evaluation of players and the adherence to team constraints result in the assembly of a well-balanced and competitive team capable of excelling in diverse match situations. This approach not only enhances the team's chances of success but also underscores the significance of data analytics in modern sports management, paving the way for more informed and strategic decision-making processes in cricket and beyond.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203013

VII. FUTURE WORK

Improve the accuracy of the predictions. This could be done by using more data, developing new machine learning algorithms, or improving the feature optimization process. Develop a system that can predict the outcome of individual matches. This would be more challenging than predicting the winning team, but it would be even more valuable to teams, coaches, and fans. Develop a system that can predict the performance of individual players. This would be useful for teams making player selection decisions, coaches developing training plans, and fans predicting who will win awards such as Man of the Match

REFERENCES

[1] K. Passim and N. Pandey, "Increased Prediction Accuracy in the Game of Cricket Using Machine Learning", Int. J. Data Min. Knowl. Manga. Process, vol. 8, no. 2, pp. 19–36, 2018
[2] H. Ahmad, A. Daud, L. Wang, H. Hong, H. Dagwood, and Y. Yang, "Prediction of Rising Stars in the Game of Cricket", IEEE Access, vol. 5, pp. 4104–4124, 2017
[3] V. Q. L. N. Hdvhu et al., "Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms", 4th International Conference on Electrical Engineering and Information & Communication Technology, pp. 500-505, 2018
[4] I. P. Wickramasinghe, "Predicting the performance of batsmen in test cricket", J. Hum. Sport Exec vol. 9, no. 4, pp. 744–751, 2014
[5] D. Bhattacharjee and D. G. Pachinko, "Analysis of Performance of Bowlers using Combined Bowling Rate", Int. J. Sport. Sci. Eng., vol. 06, no. 03, pp. 184–192, 2012
[6] I. P. Wickramasinghe, "Predicting the performance of batsmen in test cricket", J. Hum. Sport Exert., vol. 9, no. 4, pp. 744–751, 2014.
[7] Kemble, A., Rao, R., Kale, A., and Sam ant, S. (2011), Selection of cricket players using Analytical hierarchy process, In
[8] Barr, G. D. I. and Kantor, B. S. (2004), A criterion for comparing and selecting batsmen in limited overs cricket, Journal of the Operational Research Society, 55(12), 1266-1274.
[9] M. J. Lakshmi and S. N. Rao, ''Auto-Play: A Data Mining Approach To Odi Cricket Simulation And Prediction,'' Soft Comput., vol. 26, no. 13, pp. 6245–6253, Jul. 2022
[10] W. Jun and Z. Liyuan, ''Cricket Score And Winning Prediction Using Data Mining'' Int. J. Comput. Intell. Syst., vol. 15, no. 1, p. 35, Dec. 2022

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com