Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Cricket Match Outcome Forecasting Using Historical Data: A Machine Learning Study on India's Test Matches

Dr. Sushilkumar Kalmegh[1*], Bhushan Padar[2]
[1*]PG Department of Computer Science & Engineering, Sant Gadge Baba Amravati University, Amravati, Maharashtra, India, sushil.kalmegh@gmail.com
[2]PG Department of Computer Science & Engineering, Sant Gadge Baba Amravati University, Amravati, Maharashtra, India, bhushanpadar@gmail.com

ARTICLE INFO

Received: 10 Oct 2024

Revised: 29 Nov 2024

Accepted: 20 Dec 2024

ABSTRACT

This study presents a data-driven investigation into the prediction of match outcomes for India's Test cricket matches spanning a historical timeline from 1932 to 2022. Leveraging a dataset comprising 561 match instances with 13 curated features including match venue, toss result, opposition team, and team form this research applies three supervised machine learning algorithms: Decision Tree, Logistic Regression, and Random Forest. The objective is to assess each model's effectiveness in forecasting binary outcomes: win or loss.

Unlike prior studies that emphasize limited datasets or contemporary formats such as T20 or ODI, this work explores the long-format Test cricket context, offering a comprehensive view of performance dynamics over nine decades. Feature engineering strategies were employed to enrich contextual variables, while cross-validation techniques ensured model robustness.

The Random Forest model demonstrated superior performance in terms of accuracy and generalization, suggesting its suitability for capturing the nonlinear complexities of Test match outcomes. This research contributes to the domain of sports analytics by offering a replicable framework for long-term performance forecasting, and highlights the evolving nature of team dynamics, strategy, and opposition patterns over time.

The findings not only support the potential of machine learning in historical sports data analysis but also serve as a foundation for strategic planning and predictive modeling in cricket and similar domains.

Keyword: Match Outcome Prediction, Indian Cricket, Historical Data Analysis, Decision Tree, Logistic Regression, Random Forest, Sports Analytics

I. INTRODUCTION

Cricket, particularly Test cricket, holds a significant cultural and historical position in India. As the longest and most traditional format of the game, Test matches reflect not only a team's technical strength but also its endurance, strategy, and adaptability over multiple days of play. Predicting outcomes in Test cricket is an inherently complex task due to the extended nature of the format, variability in pitch conditions, player form, opposition quality, and countless contextual factors that evolve over time. Despite its complexity, outcome forecasting in cricket has gained traction with the rise of sports analytics and machine learning technologies.

Over the past decade, several studies have explored predictive modeling in cricket using machine learning, with a dominant focus on limited-overs formats such as One Day Internationals (ODIs) and Twenty20 (T20). These formats offer more immediate and structured data, leading to easier modeling and forecasting. In contrast, Test cricket remains relatively underexplored in the domain of predictive analytics, largely due to its temporal depth, irregular match intervals, and strategic intricacies that defy short-term statistical modeling.

This research addresses this gap by investigating the feasibility of forecasting Test match outcomes for the Indian cricket team using supervised machine learning techniques. We compile a comprehensive dataset spanning from 1932 to 2022, encompassing 561 Test matches played by India. The dataset includes 13 features, such as toss result, venue type, opposition team, previous match outcomes, and more, carefully selected to represent influential match-level and contextual variables.

Three widely used and interpretable machine learning algorithms—Decision Tree, Logistic Regression, and Random Forest—are employed to develop classification models aimed at predicting binary outcomes: win or

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

loss. The study compares their performance to determine which algorithm best captures the decision-making patterns embedded in historical cricket data.

This research is novel in two key aspects. First, it applies modern machine learning methods to the traditional and complex format of Test cricket, spanning nearly a century of play. Second, it evaluates the effectiveness of classical ML algorithms in modeling such long-range, historical data, offering insights into both the game's evolution and algorithmic suitability for temporal sports data.

The findings of this study contribute to the broader field of sports analytics and demonstrate the potential of data-driven forecasting in cricket. The results may be valuable to coaches, analysts, and cricket boards for match preparation, retrospective analysis, and strategy formulation.

II. LITERATURE REVIEW

A recent study applied an Improved Grey Wolf Optimizer (GWO) to enhance the performance of a Gradient Boosting Decision Tree (GBDT) model for predicting college student sports performance, addressing key limitations of the original GWO such as premature convergence and limited global search ability. By integrating a Circle chaotic sequence for better population initialization, an adaptive position update strategy for improved exploration, and a reverse search mechanism to prevent local optima entrapment, the enhanced GWO significantly improved the prediction accuracy of the GBDT model. Although this work was applied in the context of individual sports performance, it presents valuable insights for broader sports analytics applications. The optimization strategies employed in this study demonstrate the potential for improving machine learning model effectiveness in domains with complex, non-linear patterns—such as Test cricket outcome prediction. These techniques highlight the importance of adaptive optimization in enhancing classical ML models, suggesting possible future directions for refining our models in cricket analytics.[1]

A recent study focusing on the Taiwan Volleyball League developed a predictive system to forecast match outcomes for both male and female teams using machine learning techniques. The researchers constructed two distinct datasets encompassing a wide range of performance metrics and applied models such as Ridge Regression, Random Forest, Gradient Boosting, and Long Short-Term Memory (LSTM) networks. To refine prediction accuracy, they incorporated LASSO regression for feature selection and performed hyperparameter optimization. The results revealed that the models, particularly those enhanced through feature selection and tuning, could effectively forecast match results and offer strategic insights for coaching staff. Although the study centers on volleyball, its methodology of building performance-specific datasets and combining classical and deep learning approaches is highly relevant to cricket analytics. It underscores the importance of model diversification and data-driven feature engineering, which aligns with the objectives of our research in using supervised learning to predict Test cricket outcomes.[2]

A recent study proposed an artificial intelligence-based method for analyzing sports training data, with a particular emphasis on the application of transfer learning in time series prediction. By leveraging a large collection of sports training data across various disciplines, the researchers designed a model that transfers learned patterns from one sport to another, thereby improving prediction accuracy and model generalization, especially in data-scarce environments. The study demonstrated that the use of transfer learning significantly enhanced predictive performance, reducing the error margin to within 5%. Comparative analyses further validated the superiority of transfer learning over traditional approaches in scenarios with limited labeled data. While the focus was on training performance across multiple sports, the methodology introduces a valuable approach for improving prediction under constraints common in long-term historical sports datasets. For our research on Indian Test cricket outcomes, this highlights the potential of integrating transfer learning or cross-format knowledge transfer in future work to enhance model robustness where data sparsity or imbalance may affect prediction reliability.[3]

A comprehensive study on the role of data analytics in sports highlights its transformative potential in enhancing player performance, team strategies, and overall game outcomes through advanced analytical approaches. By integrating techniques such as machine learning, statistical modeling, computer vision, and wearable sensor data, the study illustrates how teams can derive actionable insights from diverse data sources—ranging from player metrics to physiological and video data. The research draws from multiple sports disciplines to present evidence of performance improvements achieved through data-driven methods, while also acknowledging practical challenges such as data quality issues, the need for cross-disciplinary expertise, and balancing analytics with human intuition. Although the study is broad in scope, it underscores the increasing importance of data analytics in modern sports and provides a foundational framework for applying similar methodologies in cricket. In the context of our research, it supports the relevance of using machine

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

learning algorithms—such as Decision Tree, Logistic Regression, and Random Forest—to analyze historical cricket data and forecast Test match outcomes, contributing to strategic decision-making in high-stakes scenarios.[4]

A study centered on team selection and player classification in the ODI format of cricket highlights the inherent complexity and unpredictability of cricket match outcomes, particularly when the competing teams exhibit closely matched skill levels. Emphasizing that cricket is fundamentally a team sport influenced by multiple dynamic factors—such as individual performances, playing conditions, and format-specific demands—the research focuses on building a prediction model that classifies players based on performance metrics to aid in optimal team selection. While the work is specific to One Day Internationals (ODIs), the principles of using data-driven models to inform team composition and improve match prediction accuracy are highly relevant to Test cricket as well. The study reinforces the importance of comprehensive performance data and contextual understanding in model development, offering valuable insight into the multidimensional nature of cricket analytics. In the context of our research, this approach supports the application of machine learning algorithms to analyze historical team-level and match-level features, enabling more reliable outcome forecasting in the longer and strategically deeper Test format.[5]

An insightful study on data-driven analysis in cricket explores the complexities of league formats, specifically the Indian Premier League (IPL), and demonstrates how historical data and statistical metrics can be used to evaluate team potential and predict match outcomes. By leveraging data from 2008 to 2022, the researchers employed Artificial Neural Networks and Random Forest algorithms to model influential factors such as required run rate, toss result, venue, and precise match location. Their approach not only achieved an impressive prediction accuracy of 99.81% but also introduced an innovative perspective on ranking teams based on past performance—independent of ongoing tournaments—which has practical applications in areas like betting markets and investment strategies. While this research is rooted in the fast-paced, data-rich IPL format, its emphasis on multi-factorial modeling and hyperparameter tuning aligns well with our work in Test cricket. Specifically, the use of historical trends and contextual game variables reinforces the value of structured feature engineering and robust model selection—principles that are central to our supervised learning-based approach to forecasting Test match outcomes for the Indian cricket team.[6]

A study focused on the 7th edition of the T20 World Cup 2020 illustrates the growing use of machine learning in forecasting sports outcomes, particularly in cricket where global anticipation around match results is immense. Using ESPN Cricinfo data, the study implemented and compared several machine learning models to predict the tournament winner, with Random Forest achieving the highest performance using a custom accuracy metric of 80.86%. The researchers concluded that Australia would win the tournament, demonstrating the potential of predictive analytics in high-stakes sports scenarios. Although this work concentrates on the T20 format and tournament-level predictions, its approach of comparing multiple ML algorithms and quantifying predictive reliability provides a methodological foundation relevant to our work. It emphasizes the importance of selecting suitable algorithms and crafting task-specific performance metrics—an aspect that strengthens our current research on applying supervised machine learning models to predict long-format Test match outcomes for the Indian cricket team using historical data.[7]

A study exploring the role of computer intelligence in analyzing and modeling cricket highlights the sport's increasing relevance to statistical and machine learning research due to its dynamic gameplay, complex rules, and growing commercial interest. The paper presents a meta-analysis of existing works focused on match outcome prediction, noting that while various methodologies have been proposed, they often lack clarity, consistency, and standard documentation. These inconsistencies pose challenges in benchmarking progress and identifying effective modeling strategies. The study emphasizes the need for a systematic evaluation of previous approaches—highlighting both strengths and limitations—to guide future research in cricket analytics. For our work on Test match prediction, this analysis reinforces the importance of methodological transparency and structured experimentation when applying machine learning techniques. It also justifies our comparative evaluation of multiple supervised learning models on a well-defined historical dataset, which seeks to fill gaps left by fragmented or under-documented prior studies in the cricket analytics domain.[8]

A recent study emphasizes the critical role of machine learning in optimizing team selection by predicting individual player performance based on a wide array of contextual factors, including opposition strength, pitch conditions, and climatic variables. The research highlights that forming a high-performing cricket team requires not only statistical analysis of player history but also consideration of dynamic environmental influences that are often overlooked in traditional models. By incorporating comprehensive data—including team dynamics and overlooked variables such as weather and field conditions—the study presents a robust framework for player performance prediction. This approach enhances the decision-making process for team

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

managers, promoting optimal team composition and chemistry. While this work is focused on player-level analytics, its methodology and insights are highly applicable to our research on Test match outcome prediction. It reinforces the importance of integrating diverse contextual features into machine learning models and justifies the inclusion of environmental and situational variables in our predictive framework for modeling India's Test cricket results using historical data.[9]
A study analyzing the Bangladesh Premier League (BPL) T20 tournament demonstrates the growing role of machine learning in cricket analytics, particularly for decision-making by investors, coaches, and analysts. By comparing five base classifiers with five ensemble classifiers using two datasets—one containing only pre-match information and the other including post-match features—the research offers a thorough evaluation of algorithm performance under different contextual conditions. Notably, K-Nearest Neighbor (KNN) achieved the highest accuracy in predicting outcomes using only pre-match data, while Gradient Boosting emerged as the most robust classifier overall. Although the study is centered on the T20 format and franchise-based league play, its systematic approach to classifier evaluation and use of pre- and post-match datasets provides valuable insights for predictive modeling in other formats. The methodology underscores the effectiveness of ensemble learning and tailored feature sets, supporting the relevance of applying similar strategies—such as Random Forests and structured match features—in our work on predicting Test match outcomes for the Indian cricket team using long-term historical data.[10]

III. METHODOLOGY

This research follows a structured and replicable machine learning pipeline comprising several key stages: data acquisition, preprocessing, feature engineering, model development, and performance evaluation. Each phase is carefully designed to ensure reliability, accuracy, and reproducibility in predicting Test match results for the Indian cricket team. An overview of the entire methodology is illustrated in Figure 1.

A. Data Collection

To build a robust prediction system, match data was collected from multiple credible sources, including official International Cricket Council (ICC) archives, the ESPN Cricinfo API, and publicly available cricket datasets on Kaggle. The dataset spans all matches played by India in the Test format from 1932 to 2022, totaling 561 instances. Each match record captures a wide range of information, such as:

- **Match-level details**: date, venue, format, and series context
- **Team information**: names of competing teams, team combinations, and historical win-loss records
- **Toss data**: which team won the toss and their decision to bat or bowl
- **Player metrics**: batting averages, strike rates, bowling economy, and player rankings
- **Environmental conditions**: weather parameters like temperature, humidity, and cloud cover, along with pitch characteristics

This diverse set of variables ensures the model is trained on both contextual and performance-related insights that influence Test match outcomes.

B. Data Preprocessing

Raw cricket data often contains inconsistencies, missing values, and categorical variables, all of which must be addressed before applying any machine learning algorithm. The following preprocessing steps were undertaken to refine the dataset:

- **Handling Missing Data**: Numerical gaps—such as unavailable player averages—were filled using median or mean values. For categorical fields like match location or toss outcome, the most frequently occurring category was used for imputation.
- **Encoding Categorical Features**: Variables like team names, venue types, and toss decisions were converted into numerical formats using label encoding and one-hot encoding techniques.
- **Scaling Numerical Values**: To ensure that features operate on comparable scales, numerical fields like player statistics and weather conditions were normalized using Min-Max scaling.
- **Removing Redundancy**: Correlation analysis and feature importance techniques were applied to identify and eliminate repetitive or non-informative variables that added noise rather than predictive value.

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

- **Dataset Splitting**: Finally, the cleaned dataset was partitioned into training and testing subsets, maintaining an 80:20 ratio. This allows for an unbiased evaluation of model performance on previously unseen data.

C. Machine Learning Models

To forecast the outcomes of India's Test cricket matches, this study employs a range of supervised machine learning algorithms. Each model is selected based on its ability to handle classification tasks and its applicability to historical sports data with mixed feature types. Below is an overview of the models used in this research:

1. Logistic Regression

Logistic Regression is a foundational classification algorithm particularly effective for binary outcome prediction, such as determining the win or loss of a cricket match. It estimates the probability of a specific outcome by analyzing input variables like historical match performance, toss results, and player statistics. The model uses a sigmoid activation function to constrain outputs between 0 and 1, providing interpretable probability scores. While it is computationally light and provides insights into feature importance, its linear nature limits its ability to capture complex, non-linear relationships often present in sports data.

2. Decision Tree Classifier

The Decision Tree algorithm structures decision-making as a tree-like model of choices and consequences. In the context of cricket, it classifies match outcomes by learning rules from training data—splitting it at each node based on the most informative features, such as toss decision, team composition, or pitch conditions. The model is straightforward to interpret and handles both categorical and numerical data types well. However, it is susceptible to overfitting, especially on noisy or small datasets. Techniques like pruning or integration into ensemble methods (e.g., Random Forest) are often used to enhance its performance.

3. Support Vector Machine (SVM)

Support Vector Machines are robust classifiers that work by finding the optimal boundary (hyperplane) that separates outcome classes—in this case, match wins or losses. With the use of kernel functions like radial basis function (RBF), SVM can model both linear and complex non-linear relationships in the data. It's especially effective on smaller datasets with well-separated classes. While SVMs are less prone to overfitting, their computational cost can become a bottleneck with large or high-dimensional datasets. Fine-tuning hyperparameters such as C and gamma is essential to achieve optimal results.

4. XGBoost (Extreme Gradient Boosting)

XGBoost is a highly efficient and scalable ensemble learning method that builds multiple decision trees in a sequential manner, where each new tree corrects the errors of its predecessor. It is particularly known for its speed and accuracy in classification problems, making it suitable for outcome prediction in cricket. Key inputs include player form, match venue, pitch conditions, and weather data. With in-built regularization mechanisms, XGBoost handles overfitting more effectively than standalone trees and can deal with missing data gracefully. With proper hyperparameter tuning, it often outperforms traditional classifiers.

5. Artificial Neural Networks (ANN)

Inspired by the structure of the human brain, Artificial Neural Networks consist of interconnected layers of neurons that process data through weighted inputs and activation functions. ANNs are particularly useful for capturing intricate, non-linear patterns in data such as those found in sports performance metrics. In this study, features like player form, environmental factors, and match context are fed into the model. While neural networks are highly accurate with sufficient training, they require large datasets, significant computational resources, and careful hyperparameter tuning. When these conditions are met, ANNs often surpass classical models in predictive accuracy.

D. Performance Evaluation

To evaluate the effectiveness of the predictive models developed in this study, four widely accepted classification metrics were used: **accuracy**, **precision**, **recall**, and **F1-score**. These metrics offer a comprehensive view of model performance by not only assessing how often predictions were correct (accuracy)

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

but also by evaluating how well the model identified positive outcomes (precision and recall) and balanced those assessments (F1-score).
A range of machine learning models—Logistic Regression, Decision Tree, Support Vector Machine (SVM), XGBoost, and Artificial Neural Network (ANN)—were assessed using the same dataset split, ensuring a fair comparison under consistent conditions. The results of this comparative analysis are summarized in Table 1.
Among all models tested, **XGBoost emerged as the most effective**, achieving the highest **accuracy of 82%**. It also recorded strong scores across the other evaluation metrics, including a **precision of 0.81**, **recall of 0.82**, and an **F1-score of 0.81**, reflecting a balanced ability to correctly classify match outcomes. **ANN followed closely**, with an accuracy of **80%**, showing competitive prediction capability and a good trade-off between precision and recall.
The **SVM model** also performed reasonably well, achieving **75% accuracy**, while the **Decision Tree** and **Logistic Regression** models trailed behind, with **72%** and **65%** accuracy respectively. These results highlight the advantages of ensemble and deep learning approaches over traditional models in capturing the nuanced patterns in historical cricket data.

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	65%	0.64	0.63	0.63
Decision Tree	72%	0.71	0.72	0.71
SVM	75%	0.74	0.75	0.74
XGBoost	82%	0.81	0.82	0.81
ANN	80%	0.79	0.80	0.79

Table 1: Performance Comparison of Machine Learning Models

IV. IMPLEMENTATION OF THE PROPOSED MODEL

The proposed cricket match prediction models were implemented using Python, leveraging a range of powerful data science libraries such as **Pandas**, **NumPy**, **scikit-learn**, **XGBoost**, and **TensorFlow**. These tools collectively facilitated efficient data handling, model development, training, evaluation, and optimization.
The process began with importing historical cricket match data—stored in CSV format—into a Pandas DataFrame. The dataset contained detailed records, including team names, individual player statistics, toss outcomes, environmental conditions, and final match results. Initial preprocessing involved addressing missing values, which are common in cricket datasets due to incomplete player stats or unrecorded weather conditions. These gaps were imputed using statistical measures like mean, median, or mode depending on the nature of the variable.
Many of the dataset's attributes—such as team names, venues, and player roles—were categorical and required transformation into numerical formats. This was accomplished using encoding techniques like **Label Encoding** and **One-Hot Encoding** to ensure compatibility with machine learning algorithms. Continuous numerical features, such as batting averages, strike rates, and weather parameters, were normalized using **Min-Max Scaling** or **Standard Scaling** to standardize the input ranges.
Following preprocessing, the dataset was split into training and testing subsets using an 80:20 ratio. This ensured that model training was based on past data, while performance evaluation was conducted on previously unseen data to assess generalizability.
Each model—Logistic Regression, Decision Tree, SVM, Random Forest, XGBoost, and ANN—was trained on the training subset. Hyperparameter tuning was carried out using **GridSearchCV** and **RandomizedSearchCV** to optimize each model's performance. The .predict() method was used to generate predictions, which were then evaluated using classification metrics like accuracy, precision, recall, and F1-score.
Among all models tested, **XGBoost achieved the highest performance**, with an accuracy of **82%**, showcasing its strength in handling structured and tabular data. **Artificial Neural Networks (ANNs)** followed with **80% accuracy**, showing high potential for learning complex patterns, albeit with higher computational requirements. **Random Forest** and **SVM** achieved **78%** and **75%** accuracy respectively, while **Decision Tree** reached **72%**. **Logistic Regression**, due to its linear limitations, lagged behind with **65% accuracy**, struggling with the dataset's non-linear complexity.

V. CONCLUSION

This research explored the use of machine learning models to predict the outcomes of India's Test cricket matches, evaluating each model's effectiveness using metrics such as accuracy, precision, recall, and F1-score. The results demonstrated that **XGBoost** and **Artificial Neural Networks (ANN)** significantly outperformed traditional models like Logistic Regression and Decision Trees by effectively capturing non-linear patterns and complex feature interactions in historical cricket data.

While **Support Vector Machines (SVM)** offered decent performance when hyperparameters were fine-tuned, challenges related to scalability and computation time were noted, especially with larger datasets. Overall, the study confirms that ensemble-based models and deep learning approaches offer superior accuracy and are better suited for real-world applications in sports analytics.

Despite promising results, the research also recognizes certain limitations, including inconsistent data availability over long historical spans and the computational cost associated with training more complex models. Future research may benefit from integrating additional real-time features such as live player form, ball-by-ball match progression, or advanced contextual indicators.

By presenting a comparative framework and demonstrating model performance on real cricket data, this study serves as a practical guide for researchers and practitioners aiming to apply machine learning in sports analytics, particularly in the domain of Test cricket prediction.

REFERENCES:

1. R. Fan and Y. Fan, "A University Sports Performance Prediction Model Based on Improved Gradient Boosting Decision Tree," *2024 20th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, Guangzhou, China, 2024, pp. 1-6, doi: 10.1109/ICNC-FSKD64080.2024.10702240.
2. H. -Y. Lin, J. -J. J. Lien and Y. -N. Chang, "Machine Learning in Sports Analytics: Volleyball Match Outcome Prediction," *2024 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, Male, Maldives, 2024, pp. 1-6, doi: 10.1109/ICECCME62383.2024.10796894.
3. S. Zelong and F. Chunlan, "Sports Training Data Analysis and Model Research Based on Artificial Intelligence," *2024 IEEE 2nd International Conference on Electrical, Automation and Computer Engineering (ICEACE)*, Changchun, China, 2024, pp. 666-669, doi: 10.1109/ICEACE63551.2024.10898641.
4. M. Sharma, V. Kaushik, M. Rathore, V. K. Yadav, P. Gangwar and A. Nibiya, "Utilizing Data Analytics in Sports Performance Optimization," *2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N)*, Greater Noida, India, 2024, pp. 1522-1526, doi: 10.1109/ICAC2N63387.2024.10895017.
5. 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.
6. S. Chakrabarty, S. Jana and P. Baral, "A Data-Driven Approach for the Prediction of Team Potential and Result During Cricket Match," *2024 2nd World Conference on Communication & Computing (WCONF)*, RAIPUR, India, 2024, pp. 1-5, doi: 10.1109/WCONF61366.2024.10691967.
7. A. Basit, M. B. Alvi, F. H. Jaskani, M. Alvi, K. H. Memon and R. A. Shah, "ICC T20 Cricket World Cup 2020 Winner Prediction Using Machine Learning Techniques," *2020 IEEE 23rd International Multitopic Conference (INMIC)*, Bahawalpur, Pakistan, 2020, pp. 1-6, doi: 10.1109/INMIC50486.2020.9318077.
8. M. M. Hatharasinghe and G. Poravi, "Data Mining and Machine Learning in Cricket Match Outcome Prediction: Missing Links," *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, Bombay, India, 2019, pp. 1-4, doi: 10.1109/I2CT45611.2019.9033698.
9. A. P. C A, Suhas, S. B J, S. M. Anvekar and U. R. C G, "Optimal Cricket Team Selection Using Machine Learning," *2024 Second International Conference on Advances in Information Technology (ICAIT)*, Chikkamagaluru, Karnataka, India, 2024, pp. 1-5, doi: 10.1109/ICAIT61638.2024.10690755.
10. M. A. Pramanik, M. M. Hasan Suzan, A. A. Biswas, M. Z. Rahman and A. Kalaiarasi, "Performance Analysis of Classification Algorithms for Outcome Prediction of T20 Cricket Tournament Matches," *2022 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 2022, pp. 01-07, doi: 10.1109/ICCCI54379.2022.9740867.