Cricket performance predictions: a comparative analysis of machine learning models for predicting cricket player’s performance in the One Day International (ODI) world cup 2023

Swamynathan Sanjaykumar 1ABCDE, Karthikeyan Udaichi 2ACE, Gowtham Rajendiran 3BCD, Marian Cretu 4BCD, Zhanneta Kozina 5BCD

1 Department of Physical Education and Sports Sciences, Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur, India
2 Department of Computing Technologies, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, India
3 Department of Computing Technologies, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, India
4 Faculty of Science, Physical Education and Informatics; University of Pitesti
5 Department of Olympic and Professional Sport, Sport Games and Tourism, H.S. Skovoroda Kharkiv National Pedagogical University, Kharkiv, Ukraine

Authors’ Contribution: A – Study design; B – Data collection; C – Statistical analysis; D – Manuscript Preparation; E – Funds Collection

DOI: https://doi.org/10.58962/HSR.2024.10.1.6-19

How to Cite

Abstract
Background and purpose
Cricket, a globally renowned bat and ball sport, is the second most popular sport worldwide. The objective of the study is to utilize machine learning algorithms to predict the performance probabilities of Indian cricket players participating in the ODI Cricket World Cup 2023. Furthermore, we aim to assess and compare the predictive precision of three machine learning models such as, Random Forest, Support Vector Regression, and XGBoost.

Materials and Methods
Data collection centered on Indian One Day International cricket statistics, encompassing matches played, batting and bowling averages, catches taken, and performance predictions. We sourced this data from reputable platforms such as ESPNcricinfo and the International Cricket Council website. Our performance prediction utilized of three machine learning models such as, Random Forest, Support Vector Regression, and XGBoost. Comparative analysis was conducted, evaluating these models through essential metrics including Mean Squared Error, Root Mean Squared Error, Mean Absolute Error, and R-squared.

Results
The comparative analysis revealed that the XGBoost model consistently outperformed the others. It exhibited lower errors with the lowest Mean Squared Error, Root Mean Squared Error, and Mean Absolute Error, signifying greater predictive accuracy. XGBoost achieved the highest R-squared value, indicating a robust relationship between predictions and actual performance probabilities. Random Forest produced satisfactory results but fell short of XGBoost’s accuracy, while Support Vector Regression displayed less accurate predictions across all metrics.

Conclusions
This research demonstrates the superior predictive ability of the XGBoost model in the performance probabilities of Indian cricket players in the ODI Cricket World Cup. The practical implications underscore the significance of data-driven insights for team selection and strategy.

Keywords: cricket, random forest, support vector regression, xgboost, performance prediction
Анотація
Санджайкумар С, Картикеян У, Раджендіран Г, Крету М., Козіна ЖЛ. Прогнози ефективності гравців в крикеті: порівняльний аналіз моделей машинного навчання для прогнозування ефективності гравця в крикет на міжнародному чемпіонаті світу One Day International (ODI) 2023

Обґрунтування і мета
Крикет, всеобічним вид спортu з битою та м'ячем, є другим за популярністю видом спортu у світі. Метою дослідження є використання алгоритмів машинного навчання для прогнозування ймовірності продуктивності індійських гравців в крикет, які беруть участь у Чемпіонаті світу з крикету ODI 2023. Крім того, ми прагнемо оцінити та порівняти точність прогнозування трьох моделей машинного навчання, таких як Random Forest., підтримка векторної регресії та XGBoost.

Матеріали та методи
Збір даних зосереджено на статистиці Індійського одноденного міжнародного крикету, що включає зіграні матчі, середні показники ватинів і боулінгів, улові і прогнози ефективності. Ми отримали ці дані з авторитетних платформ, таких як ESPNcricinfo та веб-сайт Міжнародної ради з крикету. Було проведено порівняльний аналіз, оцінюючи ці моделі за допомогою основних показників, включаючи середню квадратичну помилку, середньоквадратичну помилку, середню абсолютною помилку та R-квадрат.

Результати
Порівняльний аналіз показав, що модель XGBoost стабільно перевершує інші. Він продемонстрував менші помилки з найніжчою середньою квадратичною помилкою, середньою квадратичною помилкою та середньою абсолютною помилкою, що означає більшу точність прогнозування. XGBoost досяг найвищого значення R-квадрат, що вказує на надійний зв'язок між прогнозами та фактичними ймовірностями продуктивності. Random Forest дав завдякінні результати, але не досягав до точності XGBoost, тоді як Support Vector Regression показав менш точки прогнози за всіма показниками.

Висновки
Це дослідження демонструє чудову працездатність моделі XGBoost щодо ймовірностей продуктивності індійських гравців у крикет на чемпіонаті світу з крикету ODI. Практичні наслідки підкреслюють важливість аналізу даних для відбору команди та стратегії.

Ключові слова: крикет, випадковий ліс, опорна векторна регресія, xgboost, прогноз продуктивності

Аннотация
Санджайкумар С, Картикеян В, Раджендира Г, Крету М, Козина ЖЛ. Прогнозы эффективности игроков в крикете: сравнительный анализ моделей машинного обучения для прогнозирования эффективности игрока в крикет на международном чемпионате мира One Day International (ODI) 2023

Обоснование и цель
Крикет, всемирно известный вид спорта с битой и мячом, является вторым по популярности видом спорта в мире. Целью исследования является использование алгоритмов машинного обучения для прогнозирования вероятности производительности индийских игроков в крикете, участвующих в Чемпионате мира по крикету ODI 2023. Кроме того, мы стремимся оценить и сравнить точность прогнозирования трех моделей машинного обучения, таких как Random Forest, поддержка векторной регрессии и XGBoost.

Материалы и методы
Сбор данных сосредоточен на статистике Индийского однодневного международного крикета, включающего сыгранные матчи, средние показатели ватинов и боулингов, уловы и прогнозы эффективности. Мы получили эти данные с авторитетных платформ, таких как ESPNcricinfo и веб-сайт Международного совета по крикету. В нашем прогнозе производительности использовались три модели машинного обучения, например случайный лес, опорная векторная регрессия и XGBoost. Проведен сравнительный анализ, оценивая эти модели с помощью основных показателей, включая среднюю квадратичную ошибку, среднеквадратическую ошибку, среднюю абсолютную ошибку и R-квадрат.

Результаты
Сравнительный анализ показал, что модель XGBoost стабильно превосходит другие. Он продемонстрировал меньшие погрешности с самой низкой среднеквадратической ошибкой, средней квадратической ошибкой и средней абсолютной ошибкой, что означает большую точность прогнозирования. XGBoost достиг наивысшего значения R-квадрат, что указывает на надежную связь между прогнозами и фактическими вероятностями производительности. Random Forest дал удовлетворительные результаты, но не достигая до точности XGBoost, в то время как Support Vector Regression показал менее точные прогнозы по всем показателям.

Выводы
Это исследование демонстрирует превосходную предполагаемую способность модели XGBoost по поводу вероятностей производительности индийских игроков в крикет на чемпионате мира по крикету ODI. Практические последствия подчеркивают значимость анализа данных для отбора команды и стратегии.

Ключевые слова: крикет, случайный лес, опорная векторная регрессия, xgboost, прогноз производительности
Introduction

Cricket, a sport with a storied history and global appeal, captivates millions worldwide. Among its many tournaments, the Cricket World Cup, organized every four years by the International Cricket Council (ICC), stands out as the pinnacle of the cricketing calendar. This event unites the world’s best cricketing nations, transcending boundaries and cultures to create an unparalleled fervor in the sporting world [1, 2]. The Cricket World Cup is known for its extensive fan following and global viewership, and it represents the ultimate achievement for both players and their respective nations. It gives cricketers a unique opportunity to showcase their talents on the grandest stage, under intense scrutiny [3]. As a result, individual player performances take on immense importance in determining a team's success [4]. Cricket enthusiasts and analysts have long recognized the significance of predicting player performance, as it has implications not only for team selection but also for informing cricketing strategies, engaging fans, and enriching the understanding of the broader community [5]. Accurate predictions serve as a tool for teams to optimize their lineups, adapt their strategy, and improve their chances of clinching the coveted World Cup trophy. [6] Cricket performance prediction is a burgeoning field within sports analytics, encompassing various aspects of player performance, including batting, bowling, and fielding statistics. They conducted a comprehensive review of cricket performance prediction, highlighting the importance of factors such as player form, pitch conditions, and opposition strength in the prediction process [7]. A research work provides a foundational reference for comprehending the complexities of forecasting cricket performance. A related study explored the application of machine learning techniques for predicting the outcomes of cricket matches [8, 9, 10, 11]. While their primary focus was match outcome prediction, this work underscores the potential of statistical and machine learning models in cricket analytics [12, 13, 14].

Traditional approaches to cricket performance prediction have often relied on statistical models and historical data [15, 16]. Linear regression, a standard method in this context, seeks to establish linear relationships between predictor variables, such as player statistics and performance indicators [17]. However, linear regression can oversimplify the intricate dynamics of cricket, leading to suboptimal predictions. Conventional methods frequently need to account for the nonlinear nature of cricket performance, shaped by the complex interplay of factors such as player form, pitch conditions, and opposition strengths [18, 19]. These limitations have spurred the exploration of advanced machine-learning techniques to enhance prediction accuracy.

Random Forest (RF), an ensemble learning method, has gained recognition in diverse domains, including cricket performance prediction. It employs multiple decision trees to make predictions and is known for its robustness and versatility in handling classification and regression tasks. In cricket performance prediction, Random Forest captures nonlinear relationships between player statistics and performance indicators, making it a powerful tool for forecasting individual player contributions [20]. Support Vector Regression (SVR) is a machine learning technique renowned for capturing intricate data relationships. It excels in regression tasks, particularly in cases where conventional linear models fall short. SVR aims to identify a hyperplane that optimally fits the data while minimizing errors. In the context of cricket performance prediction, SVR accommodates complex dependencies between variables like batting averages, bowling averages, and catches [21]. XGBoost algorithm, has gained popularity across various domains due to its predictive accuracy and efficiency. It excels at handling large datasets and intricate feature interactions. In cricket performance prediction, XGBoost can reveal nuanced patterns in player data, providing highly accurate forecasts [22]. By integrating these machine learning techniques, researchers aim to overcome the limitations of traditional methods and enhance the accuracy of cricket performance prediction within the context of the Indian Cricket World Cup squad.

Hypothesis. This study hypothesizes that the machine learning techniques, specifically the Random Forest, Support Vector Regression (SVR) and XGBoost models, can be effectively employed to predict the performance probabilities of Indian cricket players in the ODI Cricket World Cup 2023.

Purpose of the study. The research study has a dual purpose. Firstly, it endeavors to assess and compare the predictive precision of three machine learning algorithms, specifically Random Forest, Support Vector Regression (SVR), and XGBoost, within the context of the ODI Cricket World Cup 2023. Secondly, it strives to harness the potential of these machine learning models to forecast the
performance probabilities of Indian players partaking in the tournament [6, 23]. This investigation will primarily focus on vital performance metrics, encompassing the number of matches played, batting averages, bowling averages, and catches taken, which serve as fundamental variables for constructing predictive models. The ultimate goal of this study is to emphasize the pivotal role of data-driven insights in augmenting decision-making processes related to team selection and strategic planning, ultimately enhancing the competitive prowess of the Indian cricket team on the global stage.

Materials and Methods

Participants

Data collection was conducted meticulously, focusing on One Day International data of the number of matches played, batting and bowling averages, catches taken, and performance predictions of the 15-member squad of the Indian team before the World Cup starts (till 22-09-2023). The data sources chosen for this study are ESPN cricinfo and the official International Cricket Council (ICC) website, both renowned for their rich and reliable cricket-related data [1, 2].

Ethical principles

Our research involved the humans, and has been provided in accordance to principles embodied in the Helsinki Declaration.

Research Design

The research design for this study employed a quantitative approach, focusing on predictive analytics in the context of Indian cricket players' performance for the ODI Cricket World Cup 2023. The study's objectives included evaluating and comparing the predictive accuracy of three machine learning algorithms: Random Forest, Support Vector Regression (SVR), and XGBoost. Data collection involved gathering comprehensive One Day International (ODI) cricket statistics for 15 members of the Indian cricket team, covering performance indicators like matches played, batting averages, bowling averages, catches taken, and performance predictions from reputable sources like ESPNcricinfo and the official International Cricket Council (ICC) website [1, 2]. To ensure the accuracy and reliability of the data, a rigorous data cleaning phase was undertaken, consisting of several essential steps. Data deduplication was carried out to eliminate redundant entries associated with performance indicators, enhancing the dataset's precision. Outlier detection played a crucial role in identifying extreme values within key metrics, such as matches played, batting and bowling averages, catches taken, and performance predictions, ensuring data integrity. Data preprocessing was crucial to ensure data accuracy and involved deduplication and the detection and removal of outliers [3, 24, 25]. The dataset was split into training and testing sets to assess model performance, and key performance indicators were selected as features for prediction. Three machine learning models were employed, and various performance metrics were used for assessment.

3. The dataset is visually represented as a boxplot in (A), which is used to identify the average value of the data, how dispersed the data is, whether skewness is present, and the presence of outliers and these outliers were removed, eradicating unusual or extreme data points and aligning the dataset with a more consistent and reliable framework, as observed in (B).

Statistical Analysis

The evaluation of the Random Forest, Support Vector Regression (SVR), and XGBoost models relies on several performance metrics, including:

Mean Squared Error (MSE): \[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

This is determined as the mean of the squared residuals between the predicted and actual values, as demonstrated in Equation (1):
In this formula, $y_{\text{predict}}$ represents the model's predicted performance, $y_{\text{actual}}$ is the observed actual performance, and $n$ represents the number of data.

Root Mean Squared Error (RMSE): RMSE, derived from the MSE, computes the average separation between predicted and actual performance values, maintaining the same units as the original data. RMSE is favored for its ability to detect significant errors and outliers, as it involves a square root operation applied to the MSE, as detailed in Equation (2):

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n} \sum (y_{\text{predict}} - y_{\text{actual}})^2}$$  \hspace{1cm} (2)

Mean Absolute Error (MAE): MAE assesses the average absolute variance between predicted and observed performance values. It is calculated as the mean of the absolute residuals between predicted and actual values, as described in Equation (3):

$$\text{MAE} = \frac{1}{n} \sum |y_{\text{predict}} - y_{\text{actual}}|$$  \hspace{1cm} (3)

R-squared Metric ($R^2$): $R^2$ assesses the goodness of fit of the models and ranges from 0 to 1. A value of 1 signifies a perfect fit, while a value of 0 represents no relationship between variables. It is calculated as follows (Equation 4):

$$R^2 = 1 - \frac{\sum (y_{\text{actual}} - y_{\text{predict}})^2}{\sum (y_{\text{actual}} - y_{\text{mean}})^2}$$  \hspace{1cm} (4)

In this equation, $y_{\text{actual}}$ denotes the observed performance values, $y_{\text{predict}}$ is the predicted performance, and $y_{\text{mean}}$ is the mean of the actual performance values.

Figure 2. The performance prediction system provides an overview of the methodological process used in predicting the performance of Indian cricket players.
Figure 3. The dataset is visually represented as a boxplot in (A), which is used to identify the average value of the data, how dispersed the data is, whether skewness is present, and the presence of outliers and these outliers were removed, eradicating unusual or extreme data points and aligning the dataset with a more consistent and reliable framework, as observed in (B)
3. The dataset is visually represented as a boxplot in (A), which is used to identify the average value of the data, how dispersed the data is, whether skewness is present, and the presence of outliers and these outliers were removed, eradicating unusual or extreme data points and aligning the dataset with a more consistent and reliable framework, as observed in (B).

Figure 4. To evaluate the predictive performance of the machine learning algorithms, the dataset was divided into separate training and testing data

Results

The results present the outcomes of the analysis regarding the predictive performance of machine learning models in estimating the contributions of Indian cricket players for the 2023 ODI Cricket World Cup. This research study evaluates three models – XG Boost, Random Forest, and Support Vector Regression – revealing varying levels of accuracy Table 4. Moreover, it provides individual performance predictions for players, where the XG Boost model projects Virat Kohli as the highest performer at 81.6%, and Shardul Thakur at 42.3%, underlining the model-specific performance variations Figure 8.

Table 1

<table>
<thead>
<tr>
<th>Random Forest (RF)</th>
<th>Mean Squared Error (MSE)</th>
<th>Root Mean Squared Error (RMSE)</th>
<th>Mean Absolute Error (MAE)</th>
<th>R-Squared (R²)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.25</td>
<td>2.29</td>
<td>1.79</td>
<td>0.953</td>
</tr>
</tbody>
</table>

The Random Forest (RF) model exhibited a Mean Squared Error (MSE) of 5.25, signifying the mean of squared variances between predicted and observed performance probabilities. Additionally, the calculation of Root Mean Squared Error (RMSE) resulted in a value of 2.29, indicating the average deviation between predicted and actual probabilities. Furthermore, the Mean Absolute Error (MAE) for the RF model was 1.79, representing the average absolute discrepancies between predicted and precise performance probabilities. The analysis unveiled an R-squared (R²) value of 0.953 for the RF model, emphasizing a strong association between the predictions and actual performance probabilities.

Figure 5. Visualizing the Performance Prediction graph of the Random Forest Regression model, the predicted values (donated in red) flow with the actual values (represented in blue color) with slight
The assessment of the Support Vector Regression (SVR) model's effectiveness in predicting the performance probabilities of Indian cricket players for the Cricket World Cup involved a meticulous examination using various performance metrics.

<table>
<thead>
<tr>
<th>Support Vector Regression (SVR)</th>
<th>Mean Squared Error (MSE)</th>
<th>Root Mean Squared Error (RMSE)</th>
<th>Mean Absolute Error (MAE)</th>
<th>R - Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.43</td>
<td>4.05</td>
<td>2.90</td>
<td>0.855</td>
</tr>
</tbody>
</table>

The Support Vector Regression (SVR) model displayed a Mean Squared Error (MSE) of 6.43, denoting the mean of squared disparities between predicted and observed performance probabilities. Additionally, the Root Mean Error (RMSE) was computed at 4.05, indicating the average discrepancy between predicted and actual probabilities. The model also registered a Mean Absolute Error (MAE) of 2.90, reflecting the mean of absolute variances between predicted and precise performance probabilities. Notably, the analysis unveiled an R-squared (R²) value of 0.855 for the SVR model, indicating a reasonably strong association between the predictions and actual performance probabilities.

The model displayed a Mean Squared Error (MSE) of 5.24, indicating the average of squared differences between predicted and actual performance probabilities. Complementing this, the Root Mean Squared Error (RMSE) at 2.28 illustrated the model's proficiency in making precise predictions, with lower values signifying heightened predictive accuracy. Moreover, the Mean Absolute Error (MAE) recorded at 1.80 denoted an average deviation of approximately 1.80 percentage points in the model's predictions from actual performance probabilities. Most notably, the analysis unveiled an exceptionally high R-squared (R²) value of 0.954, indicative of a robust association between predicted and exact performance probabilities. In summary, this predictive model's exceptional performance, as depicted by these metrics, underscores its efficacy in the performance probabilities of Indian cricket players, particularly highlighted by the solid R² value of 0.954, affirming its accuracy and reliability.

The XGBoost model demonstrated the lowest MSE of 5.24, implying the highest degree of predictive accuracy. The Random Forest model also performed commendably, exhibiting a slightly higher MSE of 5.25. In stark contrast, the Support Vector Regression model yielded the highest MSE at 6.43, indicating a notable decrease in the accuracy of performance prediction. The Root Mean Squared Error (RMSE) serves as a gauge for the average difference between predicted and actual probabilities. In alignment with the MSE results, the XGBoost and Random Forest models returned lower RMSE values (2.28 and 2.29, respectively), denoting superior predictive accuracy.

Conversely, the Support Vector Regression model registered a substantially higher RMSE of 4.05, indicative of a more substantial margin of error in its predictions. The Mean Absolute Error (MAE) calculates the average of the absolute deviations between predicted and actual performance probabilities. Corresponding to the other metrics, both the XGBoost and Random Forest models portrayed lower MAE values (1.80 and 1.79, respectively), signifying a minor average variance from actual performance. In stark contrast, the Support Vector Regression model produced a notably higher MAE of 2.90, highlighting a more considerable divergence from actual performance. The R-squared (R²) metric scrutinizes the goodness of fit and the connection between predicted and actual performance probabilities. Here, the XGBoost model secured the highest R² at 0.954, suggesting an exceptionally robust relationship between predictions and actual performance. Random Forest closely followed with an R² of 0.953, affirming a closely comparable level of accuracy. In contrast, the Support Vector Regression model presented a lower R² of 0.855, signifying a somewhat diminished representation of player performance.
Figure 6. Visualizing the Performance Prediction graph of the Support Vector Regression (SVR) Regression model, the predicted values (donated in orange color) flow with the actual values (represented in blue color) with slight deviation.

The utilization of the XGBoost (XGB) model to the performance probabilities of Indian cricket players in the Cricket World Cup underwent a comprehensive evaluation, revealing its predictive process.

<table>
<thead>
<tr>
<th>XG Boost (XGB)</th>
<th>Mean Squared Error (MSE)</th>
<th>Root Mean Squared Error (RMSE)</th>
<th>Mean Absolute Error (MAE)</th>
<th>R - Squared (R²)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.24</td>
<td>2.28</td>
<td>1.80</td>
<td>0.954</td>
</tr>
</tbody>
</table>

Figure 7. Visualizing the Performance Prediction graph of the XGBoost (XGB) Regression model the predicted values (donated in chocolate color) flow with the actual values (represented in blue color) with slight deviation.
Table 4

A comparative analysis of three distinct predictive models: Random Forest (RF), Support Vector Regression (SVR), and XGBoost (XGB)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Mean Squared Error (MSE)</th>
<th>Root Mean Squared Error (RMSE)</th>
<th>Mean Absolute Error (MAE)</th>
<th>R – Squared (R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest (RF)</td>
<td>5.25</td>
<td>2.29</td>
<td>1.79</td>
<td>0.953</td>
</tr>
<tr>
<td>Support Vector Regression (SVR)</td>
<td>6.43</td>
<td>4.05</td>
<td>2.90</td>
<td>0.855</td>
</tr>
<tr>
<td>XG Boost (XGB)</td>
<td>5.24</td>
<td>2.28</td>
<td>1.80</td>
<td>0.954</td>
</tr>
</tbody>
</table>

Figure 8. Performance Prediction (Percentage) of the Indian cricket players*

*Comparative analysis of predictive algorithm models shows that the three algorithms are performed reasonably: XGBoost 95.4%, Random Forest 95.3%, and Support Vector Regression 85.5% accuracy, respectively. Based on these algorithms, Virat Kohli shows the highest expected performance at 81.6%, while Shardul Thakur has a lower prediction at 42.3%, highlighting performance variations.

**Discussion**

In this study, we conducted a thorough comparison of three machine learning models such as, Random Forest (RF), Support Vector Regression (SVR), and XGBoost (XGB), to predict the performance probabilities of Indian cricket players in the ODI Cricket World Cup 2023 [5, 6]. The results highlighted significant differences in their predictive accuracy. The most significant finding was that the XGBoost model consistently outperformed the other models. It demonstrated the lowest Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), and the highest R-squared (R^2) value, indicating the strongest association between predictions and actual performance. These results align with the hypothesis that machine learning techniques offer improved predictive accuracy compared to traditional methods [8]. This finding aligns with the growing body of literature that emphasizes the superiority of machine learning models in sports analytics. These models can capture complex, non-linear relationships in player performance data that traditional linear models often overlook. The superior accuracy of the XGBoost model has practical implications for team selection and strategy in the ODI Cricket World Cup [9, 20].

As part of the comparative analysis, we made individual performance predictions for players. The XGBoost model projected Virat Kohli’s performance at 81.6% and Shardul Thakur’s at 42.3%, showcasing variations in player expectations. These individual
predictions provide actionable insights for team managers and selectors. They can make data-driven decisions when choosing players, improving the team's competitiveness on the world stage [5]. While previous research has primarily focused on team outcomes, this study contributes by demonstrating the application of machine learning to predict individual player performances, enriching the field of cricket analytics [6, 26, 27]. For the Indian cricket team, this means that they can optimize their squad selection based on quantitative predictions. They can choose players who are expected to deliver top-notch performances, increasing their chances of success in the ODI Cricket World Cup.

It's important to acknowledge the limitations of this study. Machine learning models are dependent on the quality and quantity of data, and cricket performance can be influenced by various unquantifiable factors such as mental state and strategy [4, 28, 29]. Additionally, the study focused on Indian cricket players, and results may not be directly transferable to players from other nations. Future research should aim to refine predictive models, incorporating more variables and potentially exploring the application of machine learning in predicting other aspects of cricket, like match outcomes. On our opinion, using machine learning models will help to calculate accurately means of developing of the athletes physical qualities [30, 31].

Conclusions

This research study leveraged machine learning techniques to predict the performance probabilities of Indian cricket players in the Men's One Day International Cricket World Cup 2023. The study found that the XGBoost model outperformed Random Forest and Support Vector Regression, showcasing superior predictive accuracy with lower errors and a strong relationship between predictions and actual performance probabilities. These findings have practical implications for team selection and strategic planning, offering data-driven insights that can enhance the Indian cricket team's competitive edge on the global stage. While the study acknowledges its limitations and the complex nature of cricket performance, it sets the stage for further research in the realm of sports analytics, highlighting the importance of integrating data-driven approaches to improve decision-making processes in the sport of cricket.

Acknowledgement

The authors are very grateful to the International Cricket Council (ICC), the Board of Control for Cricket in India (BCCI), and ESPN Cricinfo.

Conflict of Interest

The authors declare that there is no Conflict of Interest.

References

2. ESPNCricinfo. [Internet]. [cited 2023 September 23]. Available from: https://www.espncricinfo.com/


Information about the authors

Swamynathan Sanjaykumar
sanjayswaminathan007@gmail.com
https://orcid.org/0000-0001-9945-2223
Department of Physical Education and Sports Sciences, Faculty of Science and Humanities,
SRM Institute of Science and Technology, Kattankulathur, India

Karthikeyan Udaichi
karthik.harris@gmail.com
https://orcid.org/0000-0003-3686-7397
Department of Computing Technologies, School of Computing, College of Engineering and Technology,
SRM Institute of Science and Technology, Kattankulathur, India

Gowtham Rajendiran
gr6047@srmist.edu.in
https://orcid.org/0000-0002-7175-0576
Department of Computing Technologies, School of Computing, College of Engineering and Technology,
SRM Institute of Science and Technology, Kattankulathur, India

Marian Cretu
https://orcid.org/0000-0003-1934-0534
cmryan_mc@yahoo.com
Faculty of Science, Physical Education and Informatics; University of Pitesti
Str. Targul din Vale, nr.1, 110040 Pitesti, Arges, Romania

Zhanneta Kozina
zhanneta.kozina@gmail.com
https://orcid.org/0000-0001-5588-4825
Department of Olympic and Professional Sport, Sport Games and Tourism,
H.S. Skovoroda Kharkiv National Pedagogical University, Kharkiv, Ukraine

Інформація про авторів

Свамінатан Санджайкумар
sanjayswaminathan007@gmail.com
https://orcid.org/0000-0001-9945-2223
Департамент фізичного виховання та спортивних наук, Факультет природничих і гуманітарних наук,
Інститут науки і технологій SRM, Каттанкулатур - 603203, Ченгалпатту, Таміл Наду, Індія

Картікеян Удаічі
karthik.harris@gmail.com
https://orcid.org/0000-0003-3686-7397
Департамент обчислювальних технологій, Школа обчислювальної техніки,
Інженерний і технологічний коледж, Інститут науки і технологій SRM,
Каттанкулатур - 603203, Ченгалпатту, Таміл Наду, Індія

Гоутам Раджендіран
gr6047@srmist.edu.in
https://orcid.org/0000-0002-7175-0576
Департамент обчислювальних технологій, Школа обчислювальної техніки,
Інженерний і технологічний коледж, Інститут науки і технологій SRM,
Каттанкулатур - 603203, Ченгалпатту, Таміл Наду, Індія
Маріан Крету
https://orcid.org/0000-0003-1934-0534
emaryan_mc@yahoo.com
Факультет науки, фізичного виховання та інформатики, Університет Питестві
Ул.Таргундін Вале, 1, 110040, Питестві, Румунія

Жаннета Козіна
zhanneta.kozina@gmail.com
https://orcid.org/0000-0001-5588-4825
Кафедра олімпійського та професійного спорту, спортивних ігор та туризму,
Харківський національний педагогічний університет імені Г.С. Сковороди, Харків, Україна

Информация об авторах

Сваминотан Санджайкумар
sanjayswaminathan007@gmail.com
https://orcid.org/0000-0003-9945-2223
Департамент фізичного воспитания и спортивных наук, Факультет естественных и гуманитарных наук,
Інститут наук и технологий SRM, Каттанкулатур - 603203, Ченгалпатта, Тамил Наду, Індія

Картикян Удаичи
karthik.harris@gmail.com
https://orcid.org/0000-0003-3686-7397
Департамент вычислительных технологий, Школа вычислительной техники,
Інженерный и технологический колледж, Інститут наук и технологий SRM,
Каттанкулатур - 603203, Ченгалпатта, Тамил Наду, Індія

Гоутам Раджендиран
gr6047@srmist.edu.in
https://orcid.org/0000-0002-7175-0576
Департамент вычислительных технологий, Школа вычислительной техники,
Інженерный и технологический колледж, Інститут наук и технологий SRM,
Каттанкулатур - 603203, Ченгалпатта, Тамил Наду, Індія

Маріан Крету
https://orcid.org/0000-0003-1934-0534
emaryan_mc@yahoo.com
Факультет науки, фізичного виховання та інформатики, Університет Питестві
Ул.Таргундін Вале, 1, 110040, Питестві, Румунія

Жаннета Козіна
zhanneta.kozina@gmail.com
https://orcid.org/0000-0001-5588-4825
Кафедра олімпійського та професійного спорту, спортивних ігор та туризму,
Харківський національний педагогічний університет імені Г.С. Сковороди, Харків, Україна

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)