International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Advancing Cricket Analytics: A Machine Learning Approach to Predicting Players andOptimizing Lineups

Kalash Rajvanshi, Hansika Sharma, Manisha Mishra, Aditya Yadav
Under the Supervision of Mr. Punit Kumar (Assistant Professor)

Department of Computer Science and Engineering
Noida Institute of Engineering and Technology, Greater Noida
U.P., India

Abstract-

Statistics have been crucial in shaping the structure and format of cricket as it has recently developed into one of the most popular international mainstream sports. Data on cricket matches and individual players are fast expanding due to novel forms like Twenty-Twenty (T-20) cricket and an increase in the number of matches over time [1]. The demand for big data analytics is increasing, as are the possibilities for using this large data successfully in many advantageous ways, such as team player selection, predicting match victors, and many more future forecasts using big data or machine learning models [2]. Without access to massive data sets or the big data platform Spark ML, machine learning linear regression models are employed to forecast team scores [3]. This thesis contains research articles and a non-statistical essay about the decision problems of when to declare during the 2nd innings of an ipl cricket match, the choice of the best team lineups in Twenty20 cricket, and an analysis of the value of fielding in cricket. The article also explores the use of machine learning techniques to forecast cricket match outcomes using historical IPL match data [4][5].

Keywords—Cricket, analytics, Data Analysis, Data Science, Machine Learning, Model Classifiers, Modelling, Prediction, Prediction Models

I. INTRODUCTION

The popularity of cricket has increased as a result of the rapid growth of sports statistical analysis, which has altered player evaluation standards and game methods. With 1.5 billion supporters throughout the world and 106 International Cricket Council member states, cricket has emerged as one of the most popular team sports. The 10 full ICC members, particularly "the big three" of India, Australia, and England, are the subject of the majority of global interest and financial attention. One-day internationals, Twenty20 cricket, and Test matches are the three main formats in which cricket is played globally [6]. Additionally, because it is the shortest and most entertaining style of the game, T20 League cricket has attracted spectators' interest. One of these is the Indian Premier League. Since its start in 2007, the Indian Premier League, one of the most well-known T20 cricket leagues in the world, has been a major success, attracting billions of dollars in investment. The Big Bash League, Bangladesh Premier League, and county cricket in England (Twenty Blast) are some other well-known leagues that spend a lot of money marketing franchise-based cricket. Every club wants to succeed and perform better, so they need a better management team, a team selection committee that chooses the best team with quality players, and a method for choosing the best batsmen based on their prior performances.

All potential variables that could influence the result of a cricket match, such as the playing surface, will be integrated into the model for the team's performance. The model for the team's performance will be based on all potential variables, including ground influences, team quality, and home-field advantage, that could have an impact on the result of a cricket match. Because the ICC rating considers the outcome, including win, draw, and defeat, as well as the success edge, wickets, and adversary rating, Nagel Kerke R2 and AIC analysis determined that these elements are

crucial. The model fitting also took toss winning into account, but it was discovered to be unimportant. The winning percentage of a team is determined by individual performances and varies depending on the field and the country. The team's lineup is largely determined by player performance. Recent match results provide insight into a batsman's form and ability to score runs at a healthy strike rate, which is necessary for Twenty20 cricket nowadays. There are various types of pitches used to play cricket, such as batting or bowling pitches, and the conditions of the pitch are also very significant.

The outcome of a match can also be impacted by bad weather, as weather factors into how well a match goes. Teams rely on players with high batting averages and steady performances in recent games because they can be critical in setting a high target score and in chasing it by handling pressure situations. Rain or any other unforeseen factors may cause a match to be called off occasionally. The Duckworth-Lewis, or D/L, system is used to reset the target in matches that are interrupted.

In one-day international matches, multiple linear regression is a useful technique to assign winning chances to the competing teams. With the help of the D-L approach, this strategy may be quickly modified to produce "in-the-run" projections.

Cricket has grown to be a multi-billion dollar industry, and any successful cricket club now relies heavily on cricket data research. The results of cricket analytics offer a deeper understanding of the players and the game, which is particularly beneficial to those who are involved in the sport, including present players, technical staff, managers, and future players. The application of statistical analysis in cricket is growing as a result of the game's quick evolution, giving players improved insights into the action.

II. Related work/ Literature Review

The papers listed in this selection cover a range of research topics related to cricket, including predicting player performance, team composition, and valuations in the Indian Premier League.

Muthuswamy and Lam (2008) discuss the use of neural networks to predict bowler performance in one-day international cricket. The study provides an interesting approach to using artificial intelligence techniques to improve player selection and decision-making [11].

Wickramasinghe (2014) focuses on predicting the performance of batsmen in test cricket, using statistical modeling to identify the most important factors that affect player performance. The study provides valuable insights into the factors that can influence performance in cricket and could inform coaching and training strategies [12].

Barr and Kantor (2004) propose a criterion for comparing and selecting batsmen in limited-overs cricket, based on their batting records. The study suggests a statistical method that could be used to inform team selection and performance analysis[13].

Iyer and Sharda (2009) explore the use of neural networks to predict athlete performance, with a focus on cricket team selection. The study provides an interesting application of artificial intelligence techniques and could inform decision-making in sports management [14].

Jhanwar and Pudi (2016) propose a team composition-based approach to predicting the outcome of one-day international cricket matches. The study suggests a statistical method that considers team composition, player performance, and other factors that could influence match outcomes [15].

Lemmer (2002) proposes the combined bowling rate as a measure of bowling performance in cricket. The study provides a statistical method for measuring bowling performance that could be useful for performance analysis and team selection [16].

Bhattacharjee and Pahinkar (2012) also propose a statistical method for analyzing bowling performance, using the combined bowling rate. The study provides a useful approach to measuring and comparing bowler performance [17].

Mukherjee (2014) uses network analysis to quantify individual performance in cricket, focusing on batsmen and bowlers. The study provides an interesting approach to performance analysis that could be useful for coaching and training [18].

Shah (2017) proposes a new performance measure for cricket, based on statistical analysis of player performance. The study provides a valuable contribution to performance analysis and player evaluation in cricket [19].

Parker, Burns, and Natarajan (2008) focus on valuations in the Indian Premier League, using econometric modeling to estimate player values. The study provides insights into the economics of cricket and could inform decision-making in sports management [20].

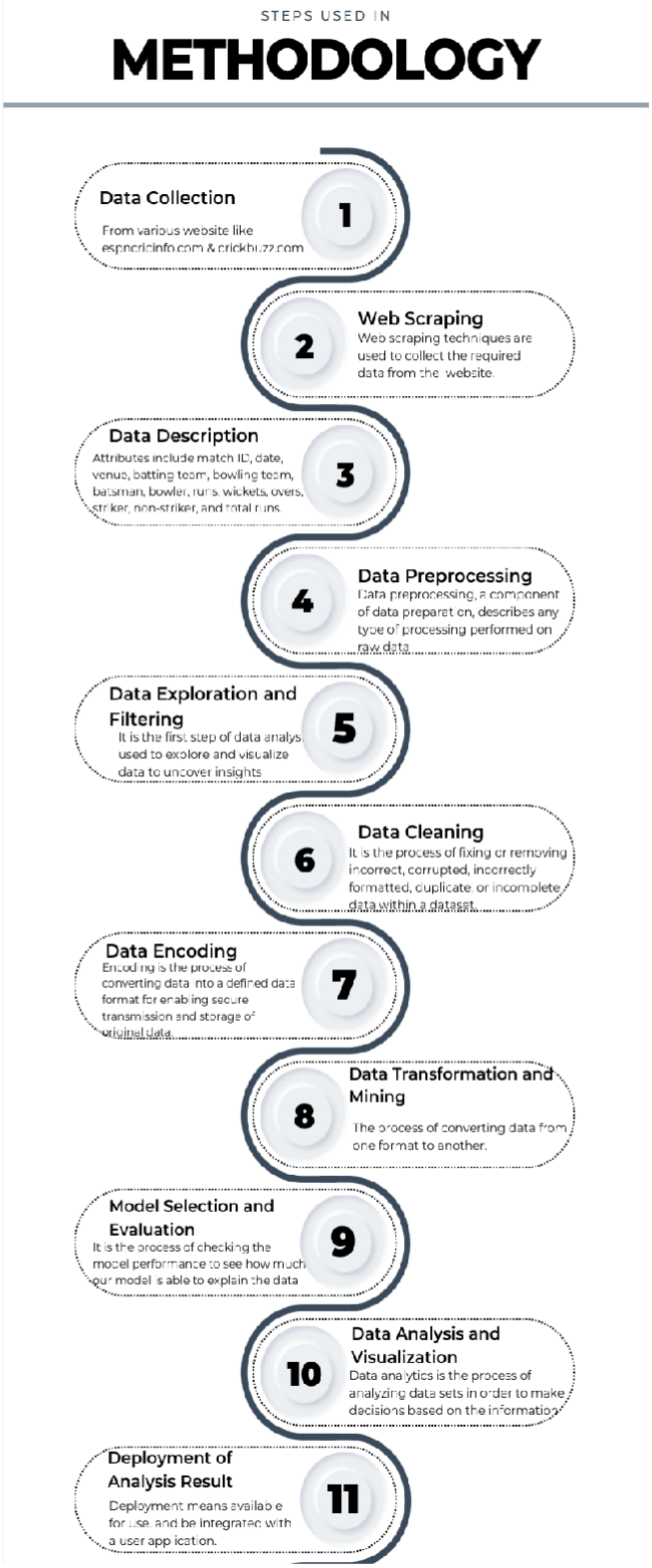
Finally, Prakash, Patvardhan, and Lakshmi (2016) propose a data analytics-based deep Mayo predictor for IPL-9. The study provides an interesting approach to using data analytics to predict outcomes in the Indian Premier League. These studies provide valuable contributions to our understanding of cricket performance analysis and could inform coaching, training, and team selection strategies.

They also highlight the potential of artificial intelligence and statistical modeling techniques for improving decision-making in sports management [21][22].

III. Methodology

The study's primary objective is to compare machine learning methods for predicting cricket matches in the IPL by building clever models considering the impact of the home-field advantage and toss winner. A pair of models have been developed based on the research, one of that includes the effect of the home field into account and the other of that to throw it into account. Pre-processing the dataset involves dividing it into two sets of attributes, removing incomplete records, and removing features that don't directly affect how well the training phase works. In order to develop predictive models for the match outcome, the research used substantial cross-validation with stratification for testing techniques and four machine learning algorithms, including the following: Naive Bayes,

Random Forest, K-nearest neighbor, and Model Decision Tree.



Data Collection:

The data is extracted from ESPN Cricinfo, which contains one-day international cricket matches from 2021 to 2022.
The dataset consists of attributes, including match details, team information, player statistics, and match outcomes.

Web Scraping:

Web scraping techniques are applied to collect the required data from the ESPN Cricinfo website.

Data Description:

The dataset contains 15 columns and 350,899 entries, representing one-day international cricket matches.
Attributes include match ID, date, venue, batting team, bowling team, batsman, bowler, runs, wickets, overs, runs_last_5, wickets_last_5, striker, non-striker, and total runs.

Data Preprocessing:

Incomplete records and data with no match result are removed from the dataset.
Irrelevant features such as match ID, date, and venue are eliminated.
Two feature sets are created: one for home ground-related features and another for toss decision-related features.

Data Exploration and Filtering:

Exploratory data analysis techniques are applied to gain insights into the dataset.
Data filtering techniques are used to identify and handle outliers, noise, and inconsistencies in the data.

Data Cleaning:

Unwanted columns, such as "first five overs," are removed from the dataset.
Only consistent teams that are valuable for the prediction are considered.
The "mid" and "date" columns are dropped from the dataset.

Data Encoding:

Categorical features, such as "bat team" and "bowl team," are converted into a one-hot encoded format using the Pandas dummies method.

Data Transformation and Mining:

Features are transformed by scaling them between 0 and 1 using the min-max scaling technique.
The min-max scaler function is applied to standardize the data.

Model Selection and Evaluation:

Different machine learning algorithms, including Naïve Bayes, Random Forest, K-nearest neighbor, and Decision Tree, are implemented on the two feature sets.
The models are evaluated using ten-fold cross-validation with stratification.
The models' performance is compared to identify the most suitable one for predicting match outcomes

Data Analysis and Visualization:

The selected model is used to analyze the dataset and predict upcoming match results.
The analysis results are visualized using appropriate charts, graphs, and statistical measures.

Deployment of Analysis Result:

The analysis results, including the selected predictive model, are deployed for forecasting upcoming IPL cricket match outcomes.

IV. CONCLUSION

In conclusion, this research paper underscores the growing significance of big data analytics, machine learning, and technological advancements in cricket. With the increasing availability of cricket-related data and the emergence of newer formats like T20 cricket, there is a growing need to leverage big data effectively for various purposes such as player selection, match outcome prediction, and performance analysis. By harnessing the power of data analysis, teams can make informed decisions regarding player selection, game strategies, and match predictions, leading to improved performance and a deeper understanding of the game. The findings presented in the paper contribute to the development of analytical tools and techniques that can enhance the overall management and success of cricket teams. In this research paper showcases the growing role of data analytics and machine learning in cricket, providing valuable insights into the game and its various facets. It highlights the potential benefits of utilizing big data techniques and advanced technologies to analyze player performance, predict match outcomes, and refine strategies. The findings of this research contribute to the broader understanding of cricket analytics and its implications for team management, coaching, and fan engagement. As the field continues to evolve, further research and exploration in cricket analytics are essential to unlocking new opportunities and enhancing the overall cricketing experience.

V. FUTURE WORK(SCOPE)

Future work related to the research presented in the paper could focus on several areas to further advance the application of data analytics and machine learning in cricket:

Integration of Real-time Data: Incorporating real-time data streams into the analysis can provide more up-to-date insights and improve the accuracy of predictions.

Player Performance Modeling: The research paper touched upon predicting player performance, but there is room for further investigation in this area. Future work could delve into more detailed modeling of player performance, considering various factors such as batting average, strike rate, bowling economy, and fielding proficiency.

Match Strategy Optimization: In future work could explore how data analytics and machine learning can inform match strategy optimization. This could involve developing algorithms to recommend optimal batting orders, bowling strategies, and fielding placements based on the analysis of historical data and real-time match situations.

Performance Evaluation Metrics: Developing robust evaluation metrics for assessing the accuracy and effectiveness of predictive models and analysis techniques in cricket is essential. Future work could focus on defining comprehensive evaluation frameworks that consider multiple factors such as model accuracy, robustness, and practical applicability.

Ethical and Privacy Considerations: As the use of data analytics and machine learning in cricket expands, ensuring ethical data handling practices and protecting player privacy becomes crucial. Future work should address these concerns and develop guidelines or frameworks to ensure responsible data usage and maintain player confidentiality.

By exploring these avenues, future research can contribute to further advancements in the application of data analytics and machine learning in cricket, improving decision-making, player performance, fan engagement, and overall game strategy.

ACKNOWLEDGMENT

We acknowledge the support of our research advisor, data contributors, colleagues, and institutions for their valuable contributions and assistance throughout this research.

REFERENCES

[1] https://en.wikipedia.org/wiki/Cricket

[2] K. Kapadia, H. Abdel-Jaber, F. Thabtah et al., Sports analytics for cricket game results using machine learning: An experimental study, Applied Computing and Informatics, https://doi.org/10.1016/j.aci.2019.11.006I.

[3] Gamage Harsha Perera, Doctor of Philosophy (Statistics), Title: *Cricket Analytics,* Examining Committee: Chair: Yi L, Associate Professor, Tim Swart: Senior Supervisor Professor, Paramjit Gil: Supervisor Associate Professor, The University of British Columbia, OkanagaBrian Naicke: Internal Examiner Director, CODE, David Stephen: External Examiner Professor, Department of Mathematics and Statistics, McGill University, 16 December 2015

[4] Mangesh Bedekar, Vinod Mane, Varad Vishwarupe (2016), Milind Pande, Uniqueness in User Behavior While Using the Web. In: Satapathy, S., Prachi M. Joshi. Proceedings of the International Congress on Information and Communication Technology. Data Analytics in the Game of Cricket: A Novel Paradigm, https://creativecommons.org/licenses/by-nc-nd/4.0

[5] Awan, M.J.; Gilani, S.A.H.; Ramzan, H.; Nobanee, H.; Yasin, A.; Zain, A.M.; Javed, R. Cricket Match Analytics Using the Big Data Approach. Electronics **2021**, 10, 2350. https://doi.org/10.3390/electronics 10192350 Academic Editors: Amin Karami, Fahimeh Jafari, and Manel Guerrero Zapata, Published: 26 September 2021

[6] https://www.icc-cricket.com/media-releases/759733

[7] Indika Wickramasinghe, Prairie View A&M University, Department of Mathematics, Prairie View, P.O. Box 519 – Mailstop 2225, TX, USA. Proceedings of the International Congress on Information and Communication Technology. Applications of Machine Learning in Cricket: A systematic review https://doi.org/10.1016/j.mlwa.2022.100435

[8] Kalpdrum Passi received his Ph.D. in Parallel Numerical Algorithms from the Indian Institute of Technology, Delhi, India in 1993. He is an Associate Professor, at the Department of Mathematics & Computer Science, at Laurentian University, Ontario, Canada.

[9] Daniel Mago Vistro, Faizan Rasheed, Leo Gertrude David "The Cricket Winner Prediction With Application Of Machine Learning And Data Analytics" in INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 8, ISSUE 09, SEPTEMBER 2019 ISSN 2277-8616

[10] https://en.wikipedia.org/wiki/Cricket_statistics

[11] S. Muthuswamy and S. S. Lam, "Bowler Performance Prediction for One-day International Cricket Using Neural Networks," in Industrial Engineering Research Conference, 2008.

[12] I. P. Wickramasinghe, "Predicting the performance of batsmen in test cricket," Journal of Human Sport & Exercise, vol. 9, no. 4, pp. 744-751, May 2014.

[13] G. D. I. Barr and B. S. Kantor, "A Criterion for Comparing and Selecting Batsmen in Limited Overs Cricket," Operational Research Society, vol. 55, no. 12, pp. 1266-1274, December 2004.

[14]] S. R. Iyer and R. Sharda, "Prediction of athletes performance using neural networks: An application in cricket team selection," Expert Systems with Applications, vol. 36, pp. 5510-5522, April 2009.

[15] M. G. Jhanwar and V. Pudi, "Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach," in European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD 2016 2016), 2016.

[16] H. H. Lemmer, "The combined bowling rate as a measure of bowling performance in cricket," South African Journal for Research in Sport, Physical Education and Recreation, vol. 24, no. 2, pp. 37-44, January 2002.

[17] D. Bhattacharjee and D. G. Pahinkar, "Analysis of Performance of Bowlers using Combined Bowling Rate," International Journal of Sports Science and Engineering, vol. 6, no. 3, pp. 1750-9823, 2012.

[18]] S. Mukherjee, "Quantifying individual performance in Cricket - A network analysis of batsmen and bowlers," Physica A: Statistical Mechanics and its Applications, vol. 393, pp. 624-637, 2014.

[19] P. Shah, "New performance measure in Cricket," ISOR Journal of Sports and Physical Education, vol. 4, no. 3, pp. 28-30, 2017.

[20] D. Parker, P. Burns, and H. Natarajan, "Player valuations in the Indian Premier League," Frontier Economics, vol. 116, October 2008.

[21] C. D. Prakash, C. Patvardhan and C. V. Lakshmi, "Data Analytics based Deep Mayo Predictor for IPL-9," International Journal of Computer Applications, vol. 152, no. 6, pp. 6-10, October 2016.

[22] M. Ovens and B. Bukiet, "A Mathematical Modelling Approach to One-Day Cricket Batting Orders," Journal of Sports Science and Medicine, vol. 5, pp. 495-502, 15 December 2006.