INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Utilizing Machine Learning for Comprehensive Analysis and Predictive Modelling of IPL-T20 Cricket Matches

Probodh Narayan Gour[1], **Mohd. Faheem Khan**[2]*

1 Department of Management, Khandelwal College of Management, Science and Technology, Bareilly, 243001, Uttar Pradesh, India
2 Department of Biotechnology, Khandelwal College of Management, Science and Technology, M.J.P. Rohilkhand University, Bareilly, 243001, Uttar Pradesh, India

OPEN ACCESS

Received: 20-11-2023

Accepted: 16-01-2024

Published: 08-02-2024

Citation: Gour PN, Khan MF (2024) Utilizing Machine Learning for Comprehensive Analysis and Predictive Modelling of IPL-T20 Cricket Matches. Indian Journal of Science and Technology 17(7): 592-597. https://doi.org/10.17485/IJST/v17i7.2944

* **Corresponding author**.

faheemphd@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objective: Current study intends to develop a predictive model for Indian Premier League (IPL) cricket match results using machine learning techniques. In order to provide a precise framework that allows for the prediction of IPL match outcomes, it aims to examine player statistics, match dynamics, and historical data. **Method:** SVM, Random Forest, Logistic Regression, Decision Tree, and KNN models were used in this study to predict player performance on any given day. Form, fitness, and previous results were among the historical player data that were used as characteristics. Each model preceded through training and testing phases, with accuracy, precision, and recall metrics evaluated to determine the most effective algorithm for forecasting player performance. **Findings:** Final studies indicated that relative team strength of competitor teams, recent form of players, and opponent pairings are distinguishing features for predicting the performance of both players and teams on any given day. The multi-machine learning approach-based model that was constructed demonstrated an accuracy of 0.71, further indicating improved performance for the given challenge. Modelling team strength is similar to modelling individual player batting and bowling performances, which is the cornerstone of our approach. **Novelty :** This paper was designed based on a novel approach leveraging combinatorial machine learning methods. This has been found to demonstrate unprecedented performance improvement in predicting a player's performance on a given day. Additionally, the presented approach may prove valuable in opening new avenues to advance machine learning applications in sports analytics by addressing the limitations of existing methods.

Keywords: Machine Learning; Sports analytics; SVM; Random Forest; KNN

1 Introduction

Recent advances in sports analytics, particularly the development of machine learning techniques, have revolutionized our understanding of sports, yielding novel insights previously unreachable. In cricket, machine learning has proved invaluable for predicting match outcomes, player performance and detecting variances in player behavioural Bias, manifested in favouritism towards players/teams or unconscious decision-making, can significantly impact cricket's fairness and integrity. This study leverages computational methods to detect and address unfairness, uncovering previously hidden patterns within extensive cricket data. Predicting cricket outcomes, like match/tournament victory, is not only desirable but necessary for safeguarding game integrity. Statistical modeling, a critical tool, utilizes diverse data points (player statistics, weather, pitch conditions, historical performance) for analysis. Algorithms like decision trees, random forests, and support vector machines are commonly employed for cricket prediction, learning from historical data to identify trends and correlations between inputs and outcomes.[1] However, predicting exact results remains challenging due to factors like complex game rules, player skill, and unpredictable elements.[2] In the Indian Premier League (IPL), eight teams engage in a round-robin format, playing each other twice (14 games per team). To predict IPL-T20 winners, we propose estimating batting and bowling potentials of the 22 players based on career statistics and recent game participation. Supervised learning methods then depict the relative superiority of one team over another using these potentials, alongside additional features like coin toss outcome and game location. This approach models batsmen, bowlers, and teams through various career statistics and recent performances.[1] The Duckworth-Lewis method calculates target scores for rain-affected limited-overs matches, but cannot forecast an individual batsman's performance.[3] Nevertheless, numerous statistical models and algorithms employed for individual player prediction consider factors like past performance, context, opponent, and other influential elements. Coaches, analysts, and fantasy cricket enthusiasts utilize these models for informed decision-making.[4] Machine learning can enhance Duckworth-Lewis prediction accuracy by learning from historical data to make more accurate target score predictions for teams batting second.[4] This increased accuracy translates to more reliable target scores for rain-affected matches. To further refine Duckworth-Lewis predictions, machine learning algorithms can be trained on past data to identify match-outcome influencing factors and leverage that knowledge for precise target score predictions for the chasing team.[5,6] Overall, machine learning holds significant potential for bolstering Duckworth-Lewis accuracy and providing dependable target scores for rain-affected matches. One key finding from the provided manuscript, not yet explored in the context of IPL T-20, is the dynamic nature of team composition. Given that a team comprises 11 players whose lineup dynamically adapts to factors like opponent, match conditions, and other variables, further investigation is warranted in this area.

2 Methodology

Our experimentation led to the identification of several key parameters, including the specifics of algorithms used to model bowlers, batsmen, and squads.[7] This section explores into the details of our approach, explicitly defining and outlining the mechanics of the chosen algorithms for modelling batsmen, bowlers, and teams. We denote the competing teams in a match m as A and B, with P(T,m) representing the set of all players from team T participating in match m, and $\varphi(p)$ signifying the set of player performance statistics for player p over their entire IPL career. Table 1, further elaborates on the principal career statistics considered for each player.[8]

Table 1. Symbolization used in algorithm

Symbolization	Details
Bet365 value	Actual performance
batting Innings	Matches in which the player batted
φBatting Average (Overall avaerage)	Matches in which the player bowled
φNum Centuries (Hundreds)	Matches played by the player
φNum Fifties (Half centuary)	Runs conceded by the player per wicket taken
φBowling Innings (Bowling Stats)	Runs scored divided by the #times the player got out
φWkts Taken (Total wkt)	Times the player has taken $\geq$ 5 wickets in a match
φFWkts Hauls	Times the player scored $\geq$ 100 runs in a match
φBowling Average (Strike Rate)	Times the player scored $\geq$ 50 but less than 100 runs in a match
φBowling Economy Average #runs conceded by the player	Wickets taken by the player

- **Dataset**

Our current study leverages a unique dataset meticulously compiled from prominent sports platforms such as BET365, Cricbuzz, and ESPN. This comprehensive assemblage encompasses detailed information on 340 matches and detailed profiles of 154 players, meticulously sourced from established and trustworthy channels. It offers rich insights into player performance and other critical game aspects, providing a valuable foundation for exploring the intricate dynamics of the sport.

- **Modelling of Batsmen**

A player's batting prowess demonstrably influences the outcome of a cricket match. In a typical eleven-player team, six to seven players specialize in batting. Comprehending a batsman's skillset relies on two key types of statistical analysis. First, we examine their career performance to gauge their overall ability. Second, we scrutinize their recent match scores to assess their current form. This latter analysis provides insight into a batsman's confidence and determines their recent contribution to the team.[9]

- **Algorithm**

The algorithm models a batsman's aptitude for a specific match through two key components: Career Score and Recent Score. Lines 2-6 calculate the batsman's Career Score, utilizing their comprehensive career statistics. Variable u (line 3) represents the ratio of matches batted to total matches played, indicating the player's specialization as a batsman. Higher values of u suggest a top-order batsman batting frequently, while lower values signify a lower-order player with fewer batting opportunities in the next match.[10] Variable φCareer Score (line 6) encapsulates all career statistics and defines the batsman's overall batting prowess. Similarly, lines 7-8 compute the batsman's Recent Score. Variable M (line 7) denotes the recent matches played, while φRecent Score (line 8) signifies the average runs scored in those games. To facilitate comparison between players with potentially diverse career and recent score ranges, these values are normalized (lines 11-12) to a common scale of [0,1]. Finally, variable φBatsman Score.[10–12]

- **Notations:** In this paper, A and B represent the two teams participating in match 'm'. We denote 'P(T, m)' as the set comprising all players in team 'T' involved in match 'm', while 'φ(p)' signifies the collection of career statistics for player 'p' as referenced in[13–16].

Algorithm used

Input: Players p∈{P(A,m)∪P(B,m)}, Career Statistics of player p:φ(p)
Output: Batsmen Score of all the players: φBatsman Score
1: for all players p ∈ {P (A, m) ∪ P (B, m)} do
2: $\varphi \leftarrow \varphi(p)$
3: u ←√ φBat Inngs φMatches Player
4: v ← 20 * φN um Centuries + 5 * φN um Fifties
5: w ← 0.3 * v + 0.7 * φBat Avg
6: φCareer Score ← u * w
7: M ← Last 4 matches played by p
8: φRecent Score ← mean(M pRuns)
9: end for
10: for all players p ∈ {P (A, m) ∪ P (B, m)} do
11: φCareer Score ← φCareer Score/max(φCareer Score :M ← Last 4 matches played by p
8: φRecent Score ← mean(M p Runs)
9: end for
10: for all players p ∈ {P (A, m) ∪ P (B, m)}
do
11: φCareer Score ← φCareer Score max(φCareer Score)
12: φRecent Score ← φRecent Score max(recent Score)
13: φBatsmen Score = 0.35 * φCareer Score + 0.65 * recent Score
14: end for

3 Results and Discussion

The burgeoning application of machine learning to predict cricket match outcomes has harvested significant attention. While inherently challenging, this pursuit presents intriguing possibilities. Machine learning's ability to analyze vast quantities of data – encompassing past matches, player statistics, weather, and pitch conditions – offers valuable insights into potential future game trajectories.[17,18]

The present work builds upon similar established principles but ventures into uncharted territory by incorporating Bet-365 betting odds, a previously untapped resource in match prediction models. Recognizing the potential influence of betting markets on match outcomes, this study explores diverse approaches for forecasting winning teams in IPL-T20 matches.[18]

The Indian Premier League itself leverages computational analysis, comparing models such as Support Vector Machines, Logistic Regression, Random Forest Regression, KNN, and Decision Trees to identify the most efficacious algorithm. Their evaluation yielded KNN as the most accurate model. However, our research demonstrates the potential of composite approaches, with an updated model incorporating positive results from multiple methods achieving superior accuracy.[19–21] This finding underscores the efficacy of transcending singular models in favour of a multi-pronged approach to this multifaceted problem.

To establish a comprehensive statistical foundation, we meticulously compiled a vast dataset from Cricbuzz and Bet-365. Encompassing all cricket matches between 2010 and 2022, this data explores into crucial match details such as participating teams, coin toss results, dates, locations, winners, and individual player performance statistics. Our primary focus targeted the top eight IPL-T-20 teams: Kolkata Knight Riders, Chennai Super Kings, Royal Challengers Bangalore, Delhi Capitals, Rajasthan Royals, Sunrisers Hyderabad, Gujarat Titans, Punjab Kings, and Lucknow Super Giants.[22]

Ensuring analytical rigor, we excluded 126 rain-affected or tied matches due to the inherent unpredictability of weather's impact on game outcomes. Furthermore, we meticulously partitioned the dataset into distinct training and test segments. The training data spans matches played between 2012 and 2015, while the test data encompasses nearly all matches from 2018 to 2020. This division, comprising a total of 340 training matches, allows for rigorous evaluation of our predictive models and analytical tools on distinct sets from different time frames, promoting the robustness and generalizability of our findings.

- **Learning Weights**

A five-year span of IPL tournaments, encompassing matches from 2014 to 2020, provided the pivotal data foundation for determining parameter weights in Algorithms A and B. To assess the influence of recent batting performances on subsequent game outcomes, we deliberately selected a sequence of consecutive matches. This enabled meticulous comparisons between players' anticipated scores, derived from the algorithms, and their actual on-field achievements. Through rigorous testing and iterative refinement of weights, we achieved significant synchronization between the top five batsmen and bowlers projected by our algorithms and the actual top five performers within each team, measured in terms of runs scored and wickets taken. This rigorous calibration process effectively validated the predictive accuracy of the algorithms, demonstrating their ability to closely align anticipated and realized performances of key players within teams.[23] Therefore, the study's methodology not only underscores the relevance of recent batting performance in shaping future outcomes but also highlights the algorithms' robust capacity to effectively forecast top performers in the IPL tournament.

- **Classifier Results**

This study comprehensively evaluated the performance of diverse binary classifiers, including SVM, Random Forests, Logistic Regression, Decision Trees, and kNN, utilizing their respective scikit-learn implementations.[13] The evaluation involved exhaustively testing all feasible parameter values and combinations for each method using the sweep feature. Notably, this rigorous process revealed the kNN algorithm with k=4 to demonstrate statistically superior efficacy compared to the best models from other classifiers, as illustrated in Figure 1.

However, predicting past match outcomes by leveraging data from future matches proved impractical. Consequently, cross-validation, with its potential to disrupt the data's chronological sequence, was deemed unfit for this purpose.

Comparing the proposed approach with previous studies, notably[10] and[11], presented an inherent challenge due to disparate underlying datasets. Our dataset lacked certain attributes crucial to the methodologies employed in these studies, such as match timing information (day/night) used in[13] and instantaneous match conditions at different phases, as utilized in[14].

Despite these disparities, the study conducted comparisons with two alternative baseline models: Model 1, assuming the team winning the toss wins the match, and Model 2, employing a positive relative strength metric (computed using Method 3)

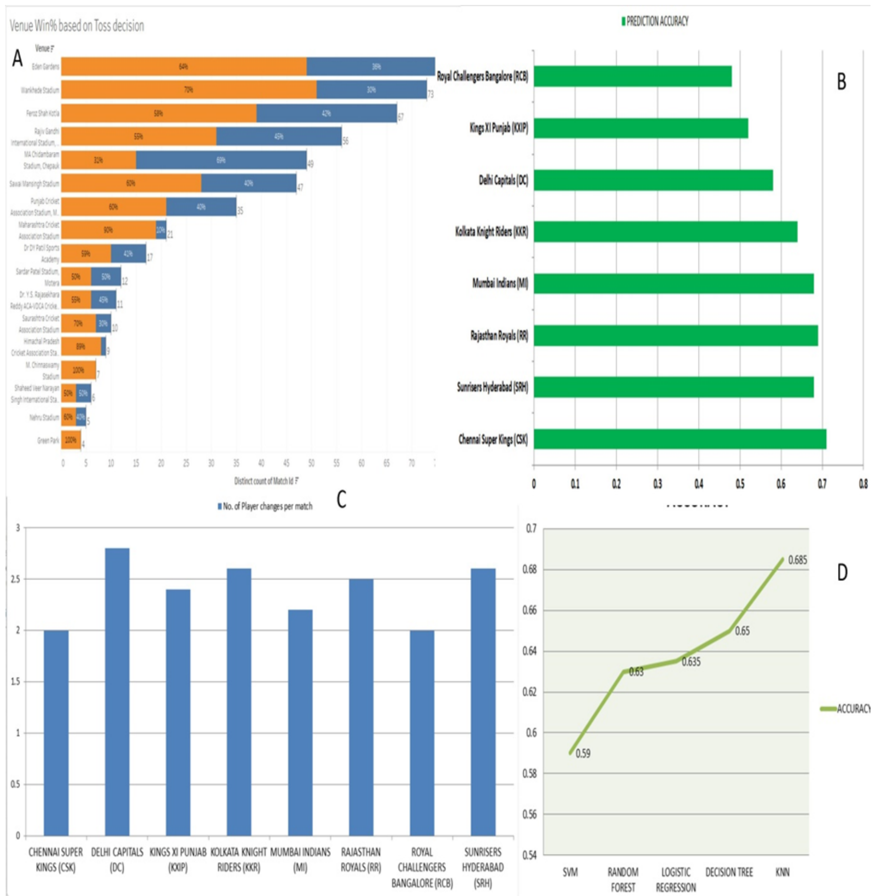

Fig 1. IPL–T20 (2015-2020) match analysis and Machine learning prediction analysis

Table 2. Comparison between implemented KNN in current study and other baseline models

MODEL	ACCURACY
MODEL A	0.54
MODEL B	0.61
OUR UPDATED MODEL	0.73

to predict the winner.[15] Detailed findings of these comparisons are presented in Table 2. Notably, our model's superiority over these rivals underscores the significance of the chosen feature combination.

In conclusion, this evaluation demonstrates the efficacy of the kNN algorithm with k=4 for binary classification within the scope of the utilized dataset. Although aligning with prior studies proved challenging due to dataset disparities, our model's marked superiority over baseline models highlights the importance of the chosen feature combination. Moving forward, reconciling dataset disparities or exploring alternative methodologies could enhance the comparability and depth of future evaluations in this domain.[16]

4 Conclusion

Leveraging statistical data from 340 IPL T20 matches, this article explores a novel approach to forecasting match outcomes that prioritizes the dynamism of player contributions. In contrast to conventional methods, this perspective treats the competitive landscape as dynamic, recognizing players as the pivotal variables impacting victory. This innovative methodology demonstrates that seemingly simple elements, when analysed through a dynamic lens, can yield compelling insights. The study's findings point towards an evolving landscape in match prediction, where player dynamics emerge as significant factors shaping game results. While KNN exhibited superior performance among individual machine learning methods assessed, a combinatorial approach leveraging consensus from all true positives yielded even higher predictive accuracy. Notably, the updated model based on this strategy achieved a noteworthy 0.73% accuracy. This finding underscores the potential of leveraging collective player dynamism insights for enhanced forecasting precision.

References

1) Duckworth FC, Lewis AJ. A fair method for resetting the target in interrupted one-day cricket matches. *Journal of the Operational Research Society*. 1998;49(3):220–227. Available from: https://doi.org/10.2307/3010471.
2) Beaudoin D, Swartz TB. The Best Batsmen and Bowlers in One-Day Cricket. *South African Statistical Journal*. 2003;37(2):203–222. Available from: https://www.sfu.ca/~tswartz/papers/bestplayers.pdf.
3) Lewis AJ. Towards fairer measures of player performance in one-day cricket. *Journal of the Operational Research Society*. 2005;56(7):804–815. Available from: https://doi.org/10.1057/palgrave.jors.2601876.
4) Swartz TB, Gill PS, Beaudoin D, deSilva BM. Optimal batting orders in one-day cricket. *Computers & Operations Research*. 2006;33(7):1939–1950. Available from: https://doi.org/10.1016/j.cor.2004.09.031.
5) Norman JM, Clarke SR. Optimal batting orders in cricket. *Journal of the Operational Research Society*. 2010;61(6):980–986. Available from: http://www.jstor.org/stable/40608270.
6) Kimber A. A Graphical Display for Comparing Bowlers in Cricket. *Teaching Statistics*. 1993;15(3):84–86. Available from: https://doi.org/10.1111/j.1467-9639.1993.tb00664.x.
7) Barr GDI, Kantor BS. A criterion for comparing and selecting batsmen in limited overs cricket. *Journal of the Operational Research Society*. 2004;55(12):1266–1274. Available from: https://doi.org/10.1057/palgrave.jors.2601800.
8) van Staden PJ. Comparison of Cricketers Bowling and Batting Performances Using Graphical Displays. *Current Science*. 2009;96(6):764–766. Available from: https://repository.up.ac.za/bitstream/handle/2263/14543/VanStaden_Comparison%282009%29.pdf.
9) Lemmer HH. The combined bowling rate as a measure of bowling performance in cricket. *South African Journal for Research in Sport, Physical Education and Recreation*. 2002;24(2):37–44. Available from: https://doi.org/10.4314/sajrs.v24i2.25839.
10) Kaluarachchi A, Varde AS. CricAI: A classification based tool to predict the outcome in ODI cricket. In: 2010 Fifth International Conference on Information and Automation for Sustainability. IEEE. 2011. Available from: https://doi.org/10.1109/ICIAFS.2010.5715668.
11) Sankaranarayanan VV, Sattar J, Lakshmanan LVS. Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction. In: Proceedings of the 2014 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics. 2014;p. 1064–1072. Available from: https://doi.org/10.1137/1.9781611973440.121.
12) Today's Cricket Match | Cricket Update | Cricket News. 2018. Available from: http://www.espncricinfo.com.
13) Croucher J. Player Ratings in One-Day Cricket. In: Cohen G, Langtry T, editors. Proceedings of the Fifth Australian Conference on Mathematics and Computers in Sport. Sydney, Australia. University of Technology. 2000;p. 95–106. Available from: https://researchers.mq.edu.au/en/publications/player-ratings-in-one-day-cricket.
14) Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn: Machine Learning Without Learning the Machinery. *GetMobile: Mobile Computing and Communications*. 2015;19(1):29–33. Available from: https://doi.org/10.1145/2786984.2786995.
15) Kavitha PV, Pavitra RS, Suwetha KP, Uvashree P. IPL Win Prediction Using Machine Learning. In: Disruptive Technologies for Big Data and Cloud Applications;vol. 905 of Lecture Notes in Electrical Engineering. Singapore. Springer. 2022;p. 457–465. Available from: https://doi.org/10.1007/978-981-19-2177-3_44.
16) Gawande K, Harale S, Pakhare S. Predictive Analysis of an Ipl Match Using Machine Learning. *International Journal of Creative Research Thoughts (IJCRT)*. 2022;10(4):795–801. Available from: https://ijcrt.org/papers/IJCRT2204215.pdf.
17) Kapadia K, Abdel-Jaber H, Thabtah F, Hadi W. Sport analytics for cricket game results using machine learning: An experimental study. *Applied Computing and Informatics*. 2022;18(3/4):256–266. Available from: https://doi.org/10.1016/j.aci.2019.11.006.
18) gajjala SR, Gonnabattula D, kanakam M, Maddipudi SS. Prediction of IPL Match Score and Winner Using Machine Learning Algorithms. *Journal of Emerging Technologies and Innovative Research*. 2021;8(6):c437–c444. Available from: http://www.jetir.org/papers/JETIR2106333.pdf.
19) Srikantaiah KC, Khetan A, Kumar B, Tolani D, Patel H. Prediction of IPL Match Outcome Using Machine Learning Techniques. In: Proceedings of the 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021);vol. 4 of Atlantis Highlights in Computer Sciences. 2021;p. 399–406. Available from: https://doi.org/10.2991/ahis.k.210913.049.
20) Awan MJ, Gilani SAH, Ramzan H, Nobanee H, Yasin A, Zain AM, et al. Cricket Match Analytics Using the Big Data Approach. *Electronics*. 2021;10(19):1–12. Available from: https://doi.org/10.3390/electronics10192350.
21) Wickramasinghe I. Applications of Machine Learning in cricket: A systematic review. *Machine Learning with Applications*. 2022;10:1–9. Available from: https://doi.org/10.1016/j.mlwa.2022.100435.
22) Bhardwaj V, Kale A, Malu S, Narsale O, Dange BJ. Prediction of IPL Winner Using Machine Learning Algorithm. *International Journal of Research Publication and Review*. 2023;4(5):5786–5789. Available from: https://ijrpr.com/uploads/V4ISSUE5/IJRPR13577.pdf.
23) Barr G, Holdsworth CG, Kantor BS. Evaluating performances at the 2007 Cricket World Cup. *South African Statistical Journal*. 2008;42(2):125–142. Available from: https://www.researchgate.net/publication/289202953_Evaluating_performances_at_the_2007_cricket_world_cup.