Vol. 3, Issue III, 2024 (October – December)
International Journal of Politics & Social Sciences Review (IJPSSR)
Website: https://ijpssr.org.pk/ OJS: https://ojs.ijpssr.org.pk/ Email: ijpssrjournal@gmail.com
ISSN 2959-6467 (Online) :: ISSN 2959-6459 (Print)
ISSN 2959-6459 (ISSN-L)
Recognized by Higher Education Commission (HEC), Government of Pakistan

Influential Factors in Successful Run Chases in T20 International Cricket: A Statistical and Predictive Analysis

Syed Asghar Ali Shah[1], Qamruz Zaman[2], Ubaid-Us-Salam Qazi[3], Touheed Iqbal[4] & Asma Mubarik[5]

[1] Department of Statistics, University of Peshawar, Pakistan Email:cricsportsresearchgroup@gmail.com
[2] Department of Statistics, University of Peshawar, Pakistan
[3] Department of Statistics, Govt. Postgraduate College, Nowshera, KP
[4] Senior Research officer, Agriculture Department (Research Wing) Peshawar
[5] Department of Mathematics, Shaheed Banazir Bhutto Woman University, Peshawar

Abstract

This study investigates the factors influencing successful run chases in T20 International cricket using data from 458 limited-overs matches. Variables such as innings scores, toss outcomes, and venue characteristics were analysed. Correlation analysis revealed significant relationships between first-innings scores and match outcomes, while the Shapiro-Wilk normality test indicated that second-innings scores deviate from normality, impacts model selection. Logistic regression identified batting order and match conditions as key predictors of run chasing success. The model demonstrated stability, evidenced by low variance inflation factors, and achieved high predictive accuracy validated by metrics including the area under the curve (AUC), precision, and F1 score. Findings indicate that a stable top order, favourable match conditions, and winning the toss significantly enhance the likelihood of successful chases. Additionally, strong middle orders and adaptive strategies contribute to overcoming high first-innings scores. Future research should focus on machine learning applications to explore complex variable interactions and the influence of psychological factors on run chases.

Keywords Factors, T20I Cricket, Accuracy, Precision, Area Under The Curve (AUC), F1 Score, Logistic Regression, Run Chases, Venue, Correlation Analysis

Introduction

Cricket has become increasingly data-driven, particularly in its limited-overs formats such as One-Day Internationals (ODIs) and Twenty20 (T20) matches. These fast-paced formats are known for their exciting moments, especially during run chases, where the team batting second attempts to surpass the target set by the opposition. Run chases are not merely a test of batting ability; they represent a complex interplay of various factors, including match conditions, pitch characteristics, toss outcomes, team rankings, and strategic decisions. While the physical skills of players remain crucial, a growing body of research highlights the significant impact of external variables such as weather, venue conditions, and toss decisions on the likelihood of a successful run chase. These complexities have prompted the use of empirical analytical methods, such as logistic regression and predictive modelling, to gain a deeper understanding of how these factors influence match outcomes. By applying these techniques, teams can make more informed decisions during matches, thereby enhancing their chances of a successful run chase. Several factors have been identified as influencing run chases in T20I cricket, including pitch conditions, match venue, toss outcomes, player positions, and the competitive strength of the teams. These factors are not independent; rather, they interact in a complex manner, making it challenging to accurately predict run chase outcomes. To address this, factorial design, a statistical approach that allows for the simultaneous study of multiple variables and their interactions, offers a robust framework for modelling and analysing run chases. This method has been widely applied across various disciplines to optimize responses and make predictions in scenarios where multiple factors contribute to the outcome. In the context of T20I cricket, factorial design can be utilized to examine the effects of variables such as match conditions, venue, toss outcomes, and toss decisions on run chases. By systematically varying these factors, researchers can identify the most influential variables and understand their combined effects on the outcome. (shah et al(2024), Shaabani, A.(2022))

The increasing availability of comprehensive match data, particularly from platforms such as ESPN Cricinfo, has facilitated in-depth analysis of key factors influencing match outcomes. This study seeks to explore and quantify the impact of variables such as first-innings score, toss outcome, pitch conditions and team rankings on the probability of a successful run chase. Statistical methods, including logistic regression, are employed to model the binary outcome of matches (win/loss) based on these factors. Through this approach, the primary drivers of success in run chases can be identified, offering valuable insights for optimizing match strategies. The findings of this research not only contribute to the growing body of cricket analytics literature but also provide data-driven recommendations for teams aiming to improve their performance when chasing targets.

Objectives of the Study

- To determine the most influential factors affecting the high run chases in T20I cricket
- To check the correlation and normality assumptions for factors affecting high run chases in T20I cricket
- To develop an optimal model for predicting highest runs in T20I cricket.

Literature Review

Over the past two decades, cricket analytics has evolved significantly, with an increasing emphasis on understanding the factors that influence match outcomes, particularly in limited-overs formats where results are more definitive. A substantial body of research has explored the impact of venue-specific conditions on run chases. De Silva and Swartz (2018) found that matches played at venues with smaller boundaries or at higher altitudes tend to yield higher run rates, which favors teams batting second. Similarly, Stern (2017) demonstrated that slower, deteriorating pitches, especially in sub continental venues, present challenges for teams attempting to chase large totals. These studies underscore the critical role venue conditions play in shaping match strategies, particularly for teams tasked with chasing targets.

The toss is a critical factor in determining the outcome of a run chase, particularly in day-night matches where environmental conditions can shift during the course of the game. Brooker and Hogan (2019) highlighted the influence of dew, especially in evening matches, which can impede bowlers' ability to control the ball. Consequently, teams that win the toss often opt to field first, anticipating more favorable batting conditions later in the match. This is consistent with the findings of Chakrabarty et al. (2020), who demonstrated that teams batting second tend to perform better in day-night matches, especially when they field first and take advantage of the evolving pitch and weather conditions. These studies collectively underscore the significance of the toss and its impact on match strategies in limited-overs cricket.

Team composition, particularly the strength of the middle order and the presence of versatile all-rounders, plays a pivotal role in the success of run chases. Wright et al. (2017) highlighted that teams with a robust middle order are better positioned to handle the pressures of a run chase, especially in situations where early wickets are lost or the required run rate increases. All-rounders provide teams with greater flexibility, offering balance in both the batting and bowling departments, which is particularly valuable during run chases when conditions can change unpredictably. Additionally, Kumar and Shah (2016) demonstrated that higher-ranked teams, with their deeper talent pools and greater experience, generally outperform others in run chases, though rankings alone are not always reliable predictors of success. Their research emphasizes that while team strength is a key factor, match-specific elements such as player form and in-game decision-making is equally critical to achieving success.

Cricket analytics has increasingly turned to statistical models for predicting match outcomes, with logistic regression being particularly useful for binary results like win/loss. Perera and Swartz (2017) employed logistic regression to forecast match results by examining various factors, such as innings scores and team rankings. Their findings indicated that higher first innings scores significantly decrease the chances of a successful chase, especially when targets surpass 300 runs in ODIs. While logistic regression is still favored in sports analytics for its simplicity and ease of interpretation, recent studies have also investigated more sophisticated models like random forests and gradient boosting. These models can better capture complex interactions among variables and enhance predictive accuracy (Baker & McHale (2018)).

Beyond quantitative factors, the psychological aspects of run chases have also been investigated. Edwards and Beech (2015) studied how pressure and momentum influence the success of run chases. Teams that establish early momentum, whether through solid partnerships or strategic acceleration, often excel in high-pressure scenarios. In contrast, teams that lose wickets early in the chase or lag behind the required run rate frequently struggle as the pressure intensifies. Consequently, psychological resilience plays a crucial role in successfully managing a run chase, especially in matches with high targets or challenging conditions.

In sports analytics, evaluating the performance of predictive models is essential, and metrics like the ROC curve and AUC have become standard methods for measuring model accuracy. Hastie, Tibshirani, and Friedman (2009) explained that ROC curves offer a visual representation of a model's capacity to differentiate between binary outcomes, while the AUC serves as a single performance metric, with values approaching 1, signifying greater accuracy. In the context of cricket analytics, these metrics are crucial for validating models such as logistic regression, ensuring they yield reliable and actionable insights for teams.

Additionally, Tripathi et al. (2020) focused exclusively on pre-match data, leveraging player career histories and team strength, and achieved a maximum prediction accuracy of 60.043% using the Random Forest algorithm. These studies highlight the competitive nature of machine learning applications in predicting outcomes in T20 cricket and emphasize the need for continuous advancements in modeling techniques. Various researchers have also adopted different methodologies to analyze match outcomes and player performances. For instance, Pathak and Wadhwa (2016) relied entirely on pre-match data, utilizing Naive Bayes, Support Vector Machines (SVM), and Random Forest to predict ODI outcomes. Similarly, Naik et al. (2022) concentrated on pre-match features, analyzing a single match based on player performance and batting order, which may not be scalable. Kumar et al. (2018) applied a Multilayer Perceptron (MLP) to pre-match data, achieving a performance rate of 57.4%. Moreover, Rahman et al. (2018) studied matches between Bangladesh and other teams from 2005 to 2017, predicting outcomes both before and after the first innings, with an SVM classifier attaining an accuracy of 63.63%.

Material and Methods

The methodology for this study focuses on examining the factors that impact successful run chases in limited-overs cricket through the application of statistical and predictive techniques.

Data Collection

The data for this study was collected from ESPN Cricinfo (2024), a trusted source for cricket statistics. The dataset encompasses information from 458 limited-overs cricket matches, including variables such as first and second innings scores, match outcomes (win/loss), toss results, venue characteristics (ground, continent), team rankings, and match/pitch conditions. These variables are crucial for analyzing the factors influencing the success of run chases. The data cleaning process involved addressing missing values, converting categorical variables such as toss outcomes and venues into factors, and ensuring overall consistency throughout the dataset. Descriptive summaries were created for key variables to gain insights into their distribution across teams and opponents.

Statistical Analysis

To examine the relationships among the variables, several statistical techniques were employed. Initially, descriptive statistics were calculated to summarize the dataset, followed by correlation analysis to identify associations between numerical variables, such as innings scores and match results. The normality of the second innings scores was assessed using the Shapiro-Wilk test, along with visualizations like QQ plots and histograms. Logistic regression was used to model the probability of winning (binary outcome: win/loss), incorporating predictors such as first innings score, toss outcome, pitch conditions, and team rankings. The dataset was split into training (70%) and testing (30%) sets to validate the model's predictive accuracy, ensuring a thorough evaluation with unseen data. Multicollinearity was examined using Variance Inflation Factor (VIF) values, and the model's performance was assessed through confusion matrices, accuracy calculations, and ROC curve analysis with AUC. The analysis was performed using R version 4.4.1.

Shapiro Wilk Test

The Shapiro-Wilk test is a statistical technique employed to evaluate the normality of the data in this study of T20I cricket. It assesses whether the distribution of the dataset concerning the factors influencing high run chases adheres to a normal distribution.

- **Null Hypothesis (H0)**: The sample data is drawn from a normally distributed population.
- **Alternative Hypothesis (H1)**: The sample data is not drawn from a normally distributed population.

A low p-value (typically less than 0.05) indicates a rejection of the null hypothesis, suggesting that the data does not follow a normal distribution. This deviation may influence the choice of subsequent statistical analyses in the study.

Logistic Regression Model

Logistic regression is a statistical technique used to model binary outcome variables by estimating the probability that an observation belongs to one of two categories based on one or more predictor variables. It utilizes the logistic function to transform linear combinations of the input features into a range between 0 and 1, enabling interpretation as probabilities (Shaabani, A.(2022)) . The model is generally represented in the following form in equation (1):

$$logit(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \text{------- (1)}$$

Where:

- p is the probability of the event occurring (e.g., success/failure).
- β_0 is the intercept.
- $\beta_1, \beta_2, ..., \beta_n$ are the coefficients for each predictor variable $X_1, X_2, ..., X_n$

Results and Discussion

Summary Statistics for Teams Average Scores and Results

The data presented in Table 1 indicates that India and the Netherlands are the most consistent teams, demonstrating high average scores in both innings and robust winning percentages (0.686 and 0.769, respectively). South Africa is notable for its impressive first-innings average of 181; however, their performance experiences a slight decline in the second innings.

In contrast, Scotland and Bangladesh face challenges with lower scores and unfavourable outcomes, revealing potential weaknesses in both innings. Overall, teams such as India, Australia, and England demonstrate balanced performance, while others like Afghanistan and Zimbabwe display greater variability between innings.

Table 1: Teams Matches Played, Average Scores and Average Results Percentage

Teams	Matches	Avg. Score First Inning	Avg. Score Second Inning	Avg. Result
Afghanistan	17	163	155	0.647
Australia	46	176	168	0.522
Bangladesh	27	169	157	0.370
England	54	176	167	0.463
Hong Kong	12	163	152	0.5
India	51	176	171	0.686
Ireland	26	173	160	0.5
Netherland	13	173	166	0.769
New Zealand	37	167	155	0.541
Pakistan	41	172	163	0.610
Scotland	20	161	141	0.4
South Africa	27	181	175	0.481
Sri Lanka	27	171	159	0.630
West Indies	43	173	160	0.442
Zimbabwe	17	168	153	0.412

Summary Statistics for Opponent Teams Average Scores and Results

Table 2 demonstrates that Australia, Bangladesh, and Ireland maintain consistent performance, boasting high average scores and strong win rates of 0.618, 0.75, and 0.733, respectively. In comparison, India and Afghanistan have lower win rates (both at 0.455), despite participating in a significant number of matches, indicating challenges in achieving victories.

Table 2: Opponent Team Matches Played, Average Score and Average Results Percentage

Opposition	Matches	Avg. Score first Inning	Avg. Score Second Inning	Avg. Result
Afghanistan	22	170	154	0.455
Australia	34	177	170	0.618
Bangladesh	16	167	160	0.75
Canada	1	172	168	0
England	41	172	157	0.366
Hong Kong	2	154	154	1
India	55	173	162	0.455
Ireland	15	174	166	0.733
Japan	1	184	184	1
Kuwait	1	153	153	1

Despite having competitive first innings scores, England has a low win percentage of 0.366. On the other hand, Hong Kong, Japan, and Kuwait have perfect win rates, although they have played very few matches, which limit the representativeness of their data. Overall, the findings indicate a diverse range of performance levels across different opposition teams.

Continent Wise Summary Statistics

Table 3 illustrates the distribution of T20I matches played across various continents. Asia has hosted the most matches, totalling 199, highlighting the region's significance in international T20 cricket.

Table 3: Continent wise Matches Played

Continents	Matches Played
Africa	63
America	54
Asia	199
Europe	68
Oceania	74

Oceania ranks second with 74 matches, while Europe and Africa have hosted 68 and 63 matches, respectively. The Americas have hosted the fewest matches, totalling 54. This data underscores the geographic concentration of T20I matches, with Asia emerging as the most active region for this format of cricket.

Summary Statistics of Factors Affecting Match Results

In Table 4, the statistical characteristics for various T20I cricket match variables are displayed based on 458 instances. The first innings (fInn) has a mean score of 172.16 and a standard deviation of 18.35, indicating considerable variability, with scores ranging from 150 to 259. In contrast, the second innings (SInn) shows a lower mean of 162.11 and a higher standard deviation of 26.26, suggesting even greater variability in scores, which range from 50 to 259.

The binary variable R reveals a nearly even win-loss distribution, with a mean of 0.53. Other variables, including batting order (BO), venue (Vn), match conditions (MC), toss outcomes (TO), pitch conditions (PC), and team rankings (RoT, Ropp), typically center around zero, indicating balanced conditions. However, they exhibit varying degrees of skewness and kurtosis, which can influence match outcomes and run-scoring biases.

Table 4: Summary Statistics for factors affecting Match Results

Var	N	Mean	SD	Med	MAD	Min	Max	Range	Skew	Kurt	SE
fInn	458	172.16	18.35	168	16.31	150	259	109	1.27	1.86	0.86
Sinn	458	162.11	26.26	162	17.79	50	259	209	-0.22	1.72	1.23
R	458	0.53	0.50	1	0.00	0	1	1	-0.12	-1.99	0.02
BO	458	-0.28	0.75	0	1.48	-1	1	2	0.50	-1.58	0.03
Vn	458	-0.05	0.84	0	1.48	-1	1	2	0.09	-1.06	0.04
MC	458	0.00	0.90	0	1.48	-1	1	2	0.00	-1.77	0.04
TO	458	0.59	0.49	1	0.00	0	1	1	-0.37	-1.87	0.02
PC	458	0.93	0.90	1	1.48	-1	2	3	-0.27	-0.97	0.04
RoT	458	0.29	0.77	0	1.48	-1	1	2	-0.55	-1.12	0.04
Ropp	458	0.27	0.76	0	1.48	-1	1	2	-0.50	-1.14	0.04

Correlation Analysis

The correlation plot in Figure 1 depicts the relationships among various variables associated with T20I cricket matches. The first innings score (fInn) exhibits a strong positive correlation with the second innings score (SInn), the match result (R), and the venue (Vn). This suggests that higher first innings scores are linked to higher second innings scores and more favorable match outcomes. Additionally, the match result (R) shows a significant positive correlation with SInn and a moderate correlation with batting order (BO) and pitch conditions (PC).

Conversely, variables such as RoT (ranking of team) and Ropp (ranking of opposing team) display a negative correlation with several others, indicating that higher rankings may be associated with lower scores in certain contexts.

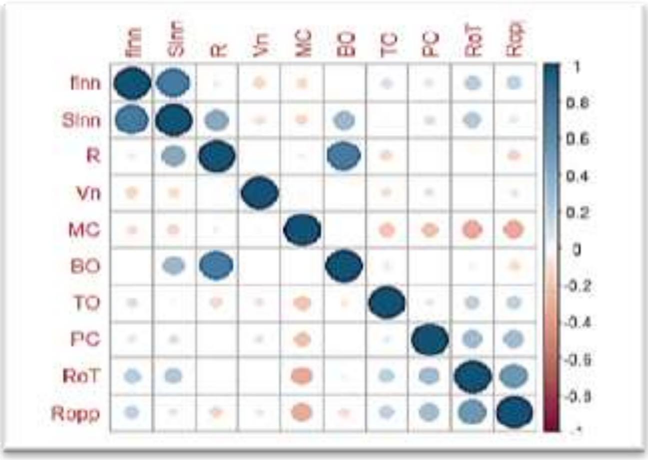

Figure 1: Correlation Plot for factors Affecting Team performance

Overall, the plot emphasizes significant interdependencies among performance metrics, match conditions, and team rankings that may impact match outcomes in T20I cricket.

Shapiro-Wilk Test for Second Inning Score

The Shapiro-Wilk test for second inning scores produced a statistic of W = 0.9733 and a P-value of 1.975×10^{-7}.

Shapiro-Wilk Normality Test	
W= 0.9733	P-value= 1.975e-07

Since the p-value is significantly lower than 0.05, we reject the null hypothesis of normality, indicating that the distribution of second inning scores is notably non-normal. This finding suggests that analyses based on the assumption of normality may not be suitable for this data.

Normal Q-Q Plot for Second Inning Score

The Q-Q plot depicted in Figure 2, illustrates how the sample quantiles of the second inning scores compare to the theoretical quantiles of a normal distribution.

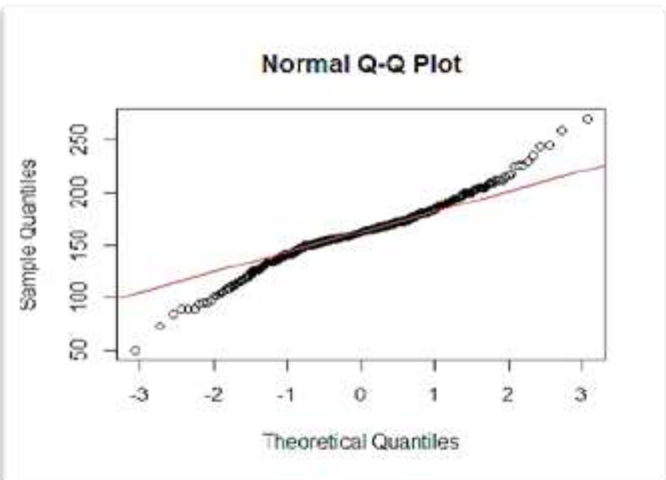

Figure 2: QQ Plot of Normality for Second Inning Score

Most data points in the plot closely follow the red reference line, indicating that the central portion of the distribution exhibits a resemblance to normality. However, there are significant deviations in the tails: the lower quantiles present a steeper curve, indicating a potential left skew, while the upper quantiles deviate upward, suggesting heavier tails than those of a normal distribution. This pattern further supports the earlier conclusion from the Shapiro-Wilk test that the second inning scores are not perfectly normally distributed.

Histogram for Second Inning Score

The histogram of second inning scores presented in Figure 3 reveals a right-skewed distribution, with a notable concentration of scores between 100 and 150. The peak of the histogram indicates that many matches resulted in scores within this range.

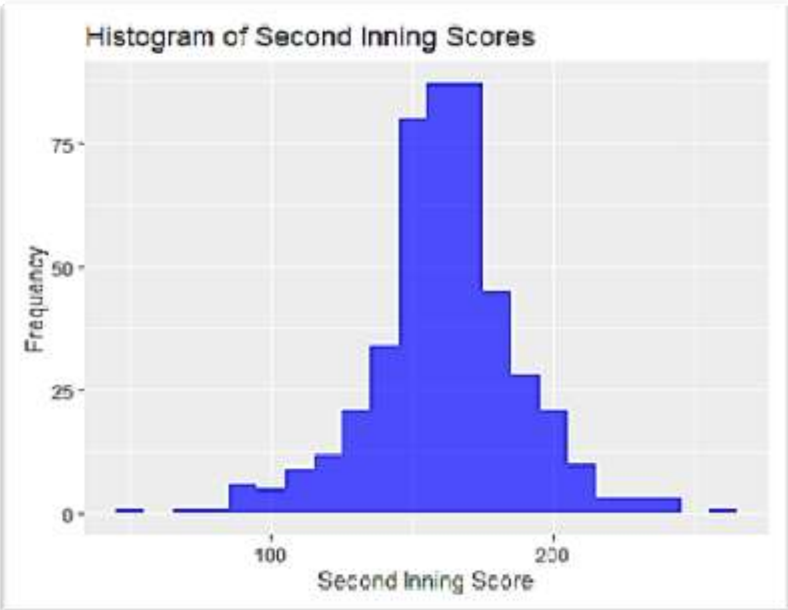

Figure 3: Histogram of Second Inning Score for T20I Matches

The distribution tapers off at higher scores, indicating fewer instances of very high scores and suggesting that high run chases are less frequent in the second innings. This visualization offers insights into T20I scoring patterns, revealing that while moderate scores are prevalent, extremely high scores are relatively uncommon.

Logistic Regression Model

In the Table 5, the results of a logistic regression analysis are displayed, assessing the significant predictors of the outcome. The intercept is not statistically significant, indicating it does not have a meaningful impact on the outcome. The first innings score (fInn) exhibits a slight positive relationship with the outcome but is not significant. The toss outcome (TO) has a negative coefficient, suggesting a decrease in the log odds of a favourable result when the toss is won, with a p-value nearing significance (0.05187). In contrast, the batting order (BO) is highly significant, with a strong positive coefficient indicating that higher batting positions significantly improve the odds of a favourable outcome.

Table 5: Summary of Logistic Regression Analysis

Var	Estimate	Std.Error	Z-Value	P- Value
Intercept	0.505650	1.539864	0.328	0.74263
fInn	0.009743	0.008882	1.097	0.27267
TO	-0.654381	0.336572	-1.944	0.05187
Vn	-0.089821	0.191523	-0.469	0.63908
BO	3.996049	0.327476	12.203	< 2e-16 ***
MC	-0.529771	0.197658	-2.680	0.00736 **
PC	0.035906	0.182629	0.197	0.84414
RoT	-0.024942	0.255201	-0.106	0.91592
Ropp	-0.418155	0.255781	-1.635	0.10209

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Match conditions (MC) also show significance, with a negative coefficient indicating their adverse effect on the outcome. In contrast, other variables such as venue (Vn), pitch conditions (PC), team ranking (RoT), and opponent ranking (Ropp) do not exhibit significant impacts on the outcome, as reflected by their high p-values. This suggests that these factors may not be influential predictors in this model.

Test of Multicollinearity

Multicollinearity occurs when two or more predictor variables in a regression model are closely correlated, sharing similar information about the variance in the outcome variable. This can lead to inflated standard errors for the coefficient estimates, making it challenging to evaluate the individual effect of each predictor. The Variance Inflation Factor (VIF) is a common method for detecting multicollinearity, with values exceeding 10 typically indicating significant multicollinearity issues.

Table 6: Test of Multicollinearity for factor affecting highest run chases

Variable	fInn	TO	Vn	BO	MC	PC	RoT	Ropp
VIF	**1.05**	**1.09**	**1.04**	**1.12**	**1.24**	**1.15**	**1.47**	**1.50**

In Table 6, the VIF values for all variables are presented, which are as follows: fInn = 1.05, TO = 1.09, Vn = 1.04, BO = 1.12, MC = 1.24, PC = 1.15, RoT = 1.47, and Ropp = 1.50. All values are significantly below the 10 threshold, indicating that there is no substantial multicollinearity. The highest VIF of 1.50 for opposition ranking (Ropp) remains low, suggesting that none of the predictor variables are highly correlated. Therefore, the logistic regression model is stable, and the coefficients can be interpreted without concerns about multicollinearity influencing the results.

Performance Metrics for Logistic Regression Model

The performance metrics for the logistic regression model demonstrate robust predictive capabilities. As shown in Table 7 and illustrated in Figure 6, the model achieves an accuracy of 0.90 and a precision of 0.91, successfully classifying 90% and 91% of instances, respectively. These results indicate a strong overall performance for the model.

Table 7: Performance Measure for Logistic Regression Model

Model	Performance Metric	Metric Value
Logistic Regression	Accuracy	0.90
	Precision	0.91
	AUC	0.94
	Sensitivity	0.89
	Specificity	0.91
	F1 Score	0.90

The Area under the Curve (AUC) is 0.94, indicating excellent discrimination between positive and negative classes; values nearing 1 reflect the model's effectiveness in distinguishing outcomes. Additionally, the sensitivity (or recall) is 0.89, meaning that 89% of actual positive cases are accurately identified, which is crucial for reducing false negatives.

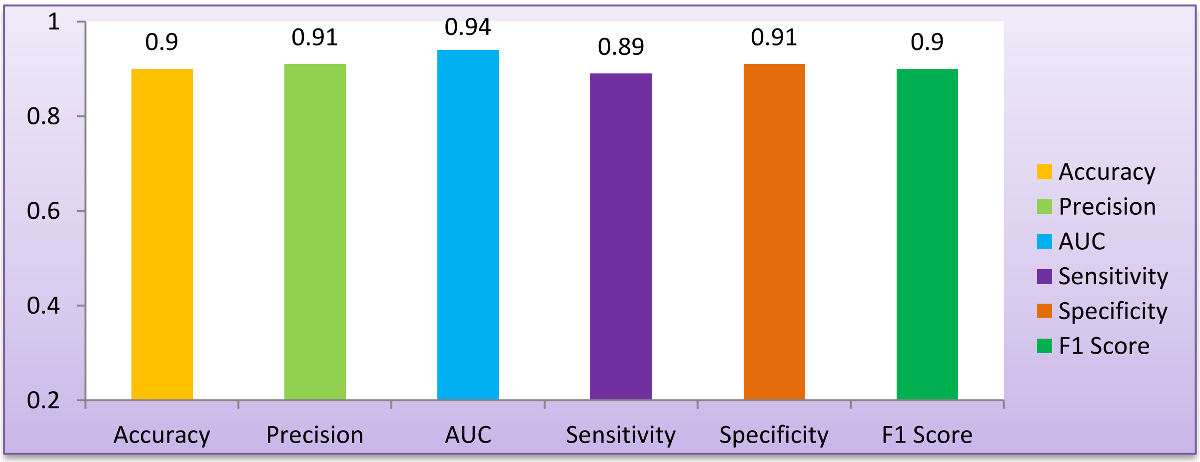

Figure 6: Bar Chart of Performance Metric for Logistic Regression Model

Furthermore, the specificity is 0.91, indicating that 91% of actual negative cases are accurately identified, demonstrating the model's ability to minimize false positives. The F1 Score of 0.90 further underscores its strong predictive accuracy. Collectively, these metrics indicate that the logistic regression model is robust and reliable for predicting outcomes in this context.

ROC Curve for Logistic Regression Model

The ROC curve displays the performance of the logistic regression model on the test set by plotting sensitivity (true positive rate) against specificity (1 - false positive rate) across different threshold settings. A curve that approaches the top-left corner signifies high sensitivity and specificity, indicating a highly effective model.

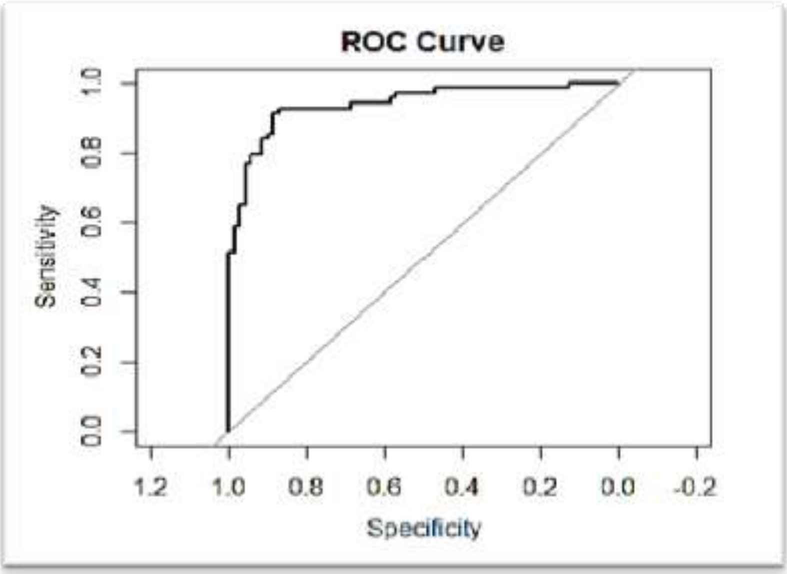

Figure 7: ROC Curve for Logistic Regression Model

The curve is notably above the diagonal line, which indicates random guessing, suggesting that the model performs significantly better than chance. The area under the curve (AUC), previously mentioned as 0.94, reflects excellent discriminatory power, indicating that the model effectively distinguishes between positive and negative classes. Overall, this ROC curve further validates the logistic regression model as a strong predictor for the outcome.

Conclusion

In conclusion, this study underscores the importance of batting order, match conditions (including venue and pitch characteristics), and toss outcomes as significant factors in successful run chases in T20 International cricket. Teams with stable top-order batsmen, favorable match conditions, and the advantage of winning the toss are more likely to achieve their chase targets. Despite the challenges posed by higher first-innings scores, teams that possess strong middle-order batsmen and employ adaptive strategies can still succeed, emphasizing the necessity of flexibility and resilience within the batting line-up.

Future research should focus on advanced predictive models, such as machine learning, to better capture the complex interplay of variables affecting run chases. Additionally, exploring psychological factors, such as pressure management in high-stakes situations, could provide valuable insights. Implementing real-time data analytics in match strategies may further enhance decision-making and improve performance in future run chases.

References

Brooker, D., & Hogan, S. (2019). The impact of dew on the outcome of day-night cricket matches. *Journal of Sports Sciences, 37*(8), 902-910 https://doi.org/10.1080/02640414.2018.1521893

Baker, R. D., & McHale, I. G. (2018). Modeling the likelihood of winning in limited-overs cricket. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 181*(1), 205-227.

Chakrabarty, A., De, A., & Saha, S. (2020). Pitch conditions and their impact on match outcomes in limited-overs cricket. *International Journal of Sports Science & Coaching, 15*(4), 620-631. https://doi.org/10.1177/1747954120935956

CricInfo. (2024). Website for cricket data. Retrieved from http://www.cricinfo.com

De Silva, B., & Swartz, T. (2018). Venue conditions and their impact on team performance in cricket. *Journal of Quantitative Analysis in Sports*, *14*(1), 45-58. https://doi.org/10. 1515/jqas-2018-0004

Edwards, S., & Beech, M. (2015). Psychological momentum and its effects on performance in limited-overs cricket. *Journal of Sports Psychology*, *7*(3), 201-210. https://doi.org/10.1080/10413200.2015.1024189.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

Kumar, J., Kumar, R., & Kumar, P. (2018). Outcome prediction of ODI cricket matches using decision trees and MLP networks. In *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)* (pp. 343-347). IEEE.

Kumar, S., & Shah, M. (2016). Team rankings and their effect on match outcomes in cricket. *International Journal of Sports Management and Marketing*, *16*(2), 123-136. https://doi.org/10.1504/IJSMM.2016.075489

Pathak, N., & Wadhwa, H. (2016). Applications of modern classification techniques to predict the outcome of ODI cricket. *Procedia Computer Science, 87*, 55-60.

Perera, H., & Swartz, T. (2017). Predicting outcomes in cricket using regression models. *Journal of Sports Analytics, 3*(2), 99-111. https://doi.org/10.3233/JSA-170021

Pramanik, M. A., Suzan, M. M. H., Biswas, A. A., Rahman, M. Z., & Kalaiarasi, A. (2022). Performance analysis of classification algorithms for outcome prediction of T20 cricket tournament matches. In *2022 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-7). IEEE.

Rahman, M. M., Shamim, M. O. F., & Ismail, S. (2018). An analysis of Bangladesh one day international cricket data: A machine learning approach. In *2018 International Conference on Innovations in Science, Engineering and Technology (ICISET)* (pp. 190-194). IEEE.

Tripathi, A., Islam, R., Khandor, V., & Murugan, V. (2020). Prediction of IPL matches using machine learning while tackling ambiguity in results. *Indian Journal of Science and Technology, 13*(38), 4013-4035.

Shah, S. A. A., Zaman, Q., Hussain, S., Iftikhar, S., Shah, S. H., Sabir, A. R., & Shah, A. (2024). Analyzing the impact of venue and match factors on ground-wise average scores and successful run chases in T-20I cricket using factorial design. *Kurdish Studies, 12*(5), 781-788.

Shaabani, A., Hamid, M., & Delavari, S. (2022). Advances in sports analytics: The integration of machine learning and design of experiments. *Journal of Sports Analytics*, *8*(3), 221-239. https://doi.org/10.3233/JSA-210463

Stern, H. (2017). The impact of pitch conditions on batting and bowling performance in international cricket. *International Journal of Performance Analysis in Sport*, *17*(2), 156-170.

Wright, M., Nelson, S., & O'Brien, J. (2017). The role of middle-order stability and finishers in successful run chases. *Cricket Analytics Review*, *12*(3), 67-75.

Zheng, S., & Man, X. (2022). An improved logistic regression method for assessing the performance of track and field sports. *Computational Intelligence and Neuroscience, 2022*, 6341495.