ACADEMIA International Journal for Social Sciences
Volume 4, Issue 3, 2025 ISSN-L (Online): 3006-6638

Quantifying the Impact of Dot Balls on Winning Probability in T20 Cricket

Muhammad Ahmad
amdm8008@gmail.com
Abdul Wali Khan University Mardan, Pakistan

Roidar Khan
roidarkhan.stats@gmail.com
University of Malakand, Pakistan

Rana Waseem Ahmad
statistics2740@gmail.com
Minhaj University Lahore, Pakistan

Fazal Wahab
wahabfazal29@gmail.com
Quaid-I-Azam University Islamabad, Pakistan

Saad Nizamani
snizamani066@gmail.com
Quaid-I-Azam University Islamabad, Pakistan

Corresponding Author: * Roidar Khan roidarkhan.stats@gmail.com

Received: 19-07-2025 **Revised:** 04-08-2025 **Accepted:** 21-08-2025 **Published:** 12-09-2025

ABSTRACT

Dot balls, defined as legal deliveries from which no runs are scored, represent a critical performance indicator in T20 cricket where efficiency on every ball determines outcomes. This study quantifies the influence of dot balls on winning probability by analyzing ball-by-ball data from 50 matches. Innings-level summaries were constructed to capture dot deliveries, total runs, wickets, and outcomes, while both statistical and machine learning models were applied to assess predictive strength. Descriptive results revealed that winning teams averaged 38 dot balls per innings compared with 45 for losing teams, reflecting a consistent performance gap. Phase-stage evaluation confirmed the section with the very best common decisive thing to be the Middle overs(7-15) in which the losers acquired almost 5 different dot deliveries in maximum cases. Looking on the outcomes of the logistic regression, it changed into located that 1 addition of dot balls lowers the probabilities of prevailing the primary vicinity through approximately thirteen percentage and on the equal time the random woodland fashions ranked dot balls as a medium however sizable predictor after the runs and wickets. The evaluation given the use of quartile in addition gave perception with the chances of win reducing strongly withinside the 2d lowest quantile constituted of seventy two percentage in least used dot balls; the pinnacle quantile of 29 percentage maximum used dot balls. These outcomes decide manage of dot-balls as a chief thing in participant fulfillment in T20 cricket in phrases of tactical importance of rotating strike and hitting a boundary.

Keywords: *T20 cricket, dot balls, rotation of strikes, prevailing the match, overall performance analytics, final results prediction, gadget learning.*

INTRODUCTION

T20 cricket has delivered approximately a revolution withinside the layout of the sport wherein it turns into a frenzied battle wherein every ball introduced can alternate the path of the sport. Contrary to different longer formats, it's far vital to be green withinside the use of all one hundred twenty prison balls and to minuscule corners cuts count. In this environment, dot balls, unproductive (runless) criminal deliveries have end up a key overall performance measure.While boundaries often capture attention, the hidden impact of dot balls on innings trajectory, pressure buildup, and eventual outcomes remains underexplored compared with conventional metrics such as total runs or wickets. Existing research in cricket analytics has examined performance determinants, win-probability models, and phase-wise strategies. Studies have consistently highlighted runs scored, wickets preserved, and strike rates as key contributors to success, yet relatively few have quantified the statistical influence of dot-ball accumulation. Early evidence suggests that excessive dot deliveries constrain scoring momentum, elevate wicket risk, and reduce win probabilities, particularly in the Middle overs when stability and strike rotation are most vital. However, most prior work either treated dot balls as secondary metrics or analyzed them descriptively without connecting them to predictive models.

This study addresses that gap by systematically quantifying the effect of dot balls on winning probability in T20 cricket. Using ball-by-ball data from 50 matches, we construct innings-level summaries, perform phase-wise analysis, and apply both logistic regression and Random Forest models to evaluate dot balls alongside runs and wickets. By integrating descriptive statistics with predictive modeling, the research establishes dot-ball control as a strategic determinant of success, offering insights relevant to players, coaches, and many recent studies have examined the determinants of T20 outcomes and the role of ball-level events; Najdan (2017) and Haworth & Mills (2024) evaluated phase-wise performance and found dot-ball-related differences between winners and losers. Several papers developed in-play or innings-level win-probability models using dynamic logistic or machine-learning approaches (Asif, 2016; Pussella, 2023; Johnstone, 2024), demonstrating that ball-by-ball features can forecast match outcomes. Work on key performance indicators for T20 has highlighted the joint importance of strike rotation (i.e., minimizing dot balls), boundary scoring, and wicket management (Kapadia, 2022; Palayangoda, 2022; Chakraborty, 2024). Recent methodological surveys and applied studies (Marshall, 2024; Jamil, 2023) emphasize the value of combining statistical models with ML algorithms to capture nonlinear effects and phase interactions, an approach echoed in several applied IPL/league analyses (multiple ML application studies; ResearchGate comps). Empirical KPI studies in women's domestic T20 and other competitions have explicitly reported that dot-ball counts and dot-ball percentage are statistically significant discriminators of match outcomes (Haworth, 2024; other KPI papers), reinforcing our focus on dot balls. Additionally, comparative work on feature importance has shown that while total runs and wickets dominate predictive power, variables capturing ball-level efficiency (dot counts, dot %, and phase-specific dots) provide complementary explanatory power (Kapadia, 2022; Najdan, 2017). Taken together, this body of work motivates a focused, phase-aware analysis of dot balls using both regression and ML tools, exactly the hybrid approach adopted in our study, which fills a gap by quantifying dot-ball effects on win probability across innings and match phases.

METHODOLOGY

Data Collection and Preparation

The dataset for this study consisted of ball-by-ball information from 50 professional T20 matches, providing a granular record of every delivery bowled. Each observation included match identifiers, innings number, batting and bowling teams, ball outcomes, runs scored, extras conceded, and wicket details. Dot balls had been outstanding as in criminal deliveries of the bat no run being brought to the rating of the batting side. On the premise of this crude statistics, summaries of innings have been constructed which protected the full wide variety of dot balls, dot-ball percentage, quantity of run scored, range of wickets misplaced and variety of deliveries faced. The end result of the suits become given with regarding the triumphing and the dropping groups accordingly forming a binary established variable (1 =

win, 0 = loss). Such prepared records enabled us to make comparisons of each losers and winners descriptively in addition to the usage of prediction models..

Variable Construction and Phase Segmentation

The overall range of dot balls in an innings became the number one unbiased variable of concern. In order to counter this, the dot-ball percent become observed that's the ratio of dot balls to prison deliveries giving a normalized degree among innings. The manipulate variables had been the full runs scored, wicket misplaced and the overall balls confronted due to the fact they may be a number of the simple determinants of the results of a healthy. Dot balls had been additionally divided into 3 vital levels to be able to get the temporal dynamics of the T20 innings: Powerplay (overs 1-6), Middle overs (7-15) and Death overs (16-20). This gave us the possibility to decide whether or not the buildup of dot-balls might make the equal effect over the direction of the innings, or whether or not there had been unique levels that have been extra harmful.

Statistical / Regression Analysis

In displaying variations among the winner and loser in descriptive information we may want to set an preliminary degree of distinction withinside the descriptive records like dot balls, runs and wickets. Correlation tables have been in the end calculated to are searching for to set up correlation amongst vast variables. In order to degree the direct impact through dot balls on results, we computed logistic regression fashions with triumphing possibility because the structured variable and dot balls, runs, wickets because the predictors. The statistical importance and realistic significance of every of the elements had been measured via way of means of odds ratios and p-values. Further level with the aid of using level logistic regressions have been conducted, wherein dot balls of the Powerplay, Middle, and Death overs have been entered independently, in order that we ought to set up which of the overs have been maximum decisive. Pseudo-R^2 information, LR-checks and class accuracy had been used to decide version fit.

Model Evaluation and Machine Learning

To in addition supplement the regression method, a Forest Classifier randomly turned into used. This version suits nicely into nonlinear relationships and consequences of interactions which might be misplaced withinside the conventional regressions. The rating of function significance became to be received with a purpose to rank the relative contribution of dot balls, runs and wickets to e.g. predictive accuracy. In order to degree the overall performance of the version, the Receiver Operating Characteristic (ROC) curves and the Area Under the Curve (AUC) cost have been carried out as nicely, evaluating the discriminative software of the logistic regression and the Random Forest. Lastly, we accomplished a opportunity evaluation in phrases of quarters, wherein we fined the quantity of dot-balls in a given innings and labeled them as both certainly considered one among 4 categories. Each quartile had the proportion of every calculated as a win periods, which converted the statistical end result of statistical findings to the intuitive insights a educate and analyst observe in practice.

RESULTS AND DISCUSSION

According to Table 1, the descriptive records in suit results have been organized among innings which resulted into victories and defeat. The maximum apparent fashion is withinside the common runs scored: the receiving facet were given 158 runs in step with inning, as compared to handiest one hundred thirty five runs acquired through the dropping aspect, a distinction of greater than 23. Dot balls additionally carry a completely vital contrast. Conquering groups used handiest a mean of 38 dot deliveries, against forty five to people who misplaced, say 7 dots balls in step with turn. The end result of this hole is extra probabilities to rating- a further boundary or greater singles, in those forms of quick codecs which could without problems paradigm shift. This has a corresponding impact at the dot-ball percent: 34.2 while the crew took a win, and simply 40.three whilst it misplaced, demonstrating that dropping facets squandered nearly in each 5 criminal deliveries. Wickets assist this meaning: Winning innings already misplaced five.five wickets, however, that's worse than the dropping innings that misplaced 7.6, which demonstrates much less strong batting and collaborations in tough comes. All those descriptive metrics blended outline a awesome connection: to lessen the wide variety of hit balls to huge partnerships and rotate the strike to the maximum, it's miles strongly related to the victory of the fits withinside the sphere of T20 cricket, whilst the criterion of excessive performance withinside the use of every ball is the pinnacle priority.

Table 1. Descriptive Statistics by Match Outcome

Is_winner	Innings	Runs	Dot_balls	Mean_dot_pct%	Mean_wickets	Balls
0.0	50.0	134.66	44.7	40.3%	7.62	111.42
1.0	50.0	158.04	38.1	34.2%	5.48	112.6

Tables 2 illustrate dot balls being given at 3 inclined ranges of T20 innings, i.e. Powerplay (overs 1-6), Middle overs (7-15), and Death overs (16-20) and outcomes cut up out among the ones having received and people having lost. There become a small, but good sized distinction (16.three vs. 17.five dot balls) withinside the variety of dot balls received via way of means of the winners and with the aid of using the losers withinside the Powerplay. The distinction is maximum major in regards to the Middle overs wherein the losers had a median of 21.eight dot deliveries in which the winners had simplest 16.7, and the distinction is extra than 5 dot balls. This distinction, thinking about the Middle overs are the longest part of play indicates that dropping groups may be efficiently crippled at some stage in the Middle overs and can not transfer strike or discover boundaries. In the Death overs, winners averaged 5.9 dot balls, while losers rose to 7.0, again highlighting inefficiency in the final push. The cumulative pattern confirms that minimizing dot deliveries across all phases is vital, but the Middle overs are particularly decisive. Teams failing to convert balls into runs in this period often enter the Death overs with insufficient momentum or too many wickets down, making late recoveries less effective. Table 2, therefore, emphasizes the strategic need to maintain tempo through the middle, where dot-ball accumulation is most damaging to winning probability.

Table 2. Dot Balls by Phase and Match Outcome

phase	is_winner	mean_phase_dot	mean_phase_runs	mean_phase_balls
Death	0	7.0	31.64	21.98
Death	1	5.95	38.0	22.26
Middle	0	21.78	66.55	58.61
Middle	1	16.66	76.46	57.46

Powerplay	0	17.48	42.86	35.52
Powerplay	1	16.32	48.9	36.0

Table 3 shows the results of logistic regression estimating the impact of batting variables on winning probability. The coefficient for dot balls is -0.14, corresponding to an odds ratio of 0.87, meaning each additional dot delivery reduces the odds of winning by approximately 13%, holding other factors constant. However, this effect is not statistically significant ($p = 0.49$) in this dataset, suggesting that dot balls alone, once runs and wickets are controlled, may not independently explain outcomes. By contrast, wickets lost exert a clear and significant effect, with a coefficient of -0.38 ($p = 0.002$), indicating that each additional wicket reduces win odds by over 30%. Total runs scored are positively associated with winning, reinforcing the intuitive expectation that higher run totals translate into victories. While dot balls lose direct significance in this multivariate framework, their indirect role is apparent: more dot balls usually suppress runs and increase pressure, raising the risk of dismissals. Thus, the regression results in Table 3 suggest that while dot balls may not always be independently decisive, they operate through their interaction with runs and wickets, reinforcing the broader message that strike rotation and scoring efficiency are central to success.

Table 3. Logistic Regression Results

Variable	Coef	StdErr	p-value
dot_balls	-0.1355	0.198	0.493
total_runs	0.0109	0.013	0.389
wickets	-0.3803	0.120	0.002

Table 4 shows the feature importance scores from the Random Forest model, which ranks the contribution of each predictor to match outcome classification. Total runs scored emerged as the strongest predictor with an importance value of 0.29, closely followed by wickets lost (0.26). Together, these two variables account for more than half of the model's predictive power, reflecting the centrality of scoring volume and batting stability. Dot balls ranked third, with an importance score of 0.21, indicating that while secondary to runs and wickets, they still make a meaningful contribution to the model's predictive accuracy. Dot-ball percentage also carried relevance (0.19), underscoring that both absolute and relative dot-ball counts matter. Total balls faced contributed the least (0.06), reflecting that innings length is fixed in T20 cricket and carries little explanatory value. The Random Forest findings reinforce the regression results by confirming that runs and wickets dominate, but they also validate dot balls as a non-trivial predictor. In combination, these variables create a comprehensive picture: runs and wickets define outcomes most strongly, while dot balls exert an indirect but important influence, especially when they accumulate excessively and constrain scoring opportunities.

Table 4. Random Forest Feature Importance

feature	importance
total_runs	0.29
wickets	0.258
dot_balls	0.207
dot_pct	0.185
total_balls	0.059

Table 5 shows the correlation matrix among the key performance indicators: dot balls, runs, wickets, dot percentage, and winning outcomes. The table highlights several critical associations. Dot balls are strongly negatively correlated with total runs (-0.36), reflecting the intuitive point that every delivery not scored from reduces a team's aggregate. Dot balls are also positively correlated with wickets (0.63), suggesting that teams struggling to rotate strike are more likely to succumb to dismissals under pressure. Importantly, dot balls are negatively associated with winning (-0.30), meaning teams that accumulate excessive dots are less likely to secure victory. In contrast, runs scored are positively correlated with winning (0.28), while wickets are negatively correlated (-0.22). Dot-ball percentage is the most closely tied to inefficiency, showing a strong negative association with runs (-0.77), confirming that a high proportion of wasted deliveries severely limits scoring momentum. Collectively, these correlations provide statistical evidence for the intuitive cricketing insight that dot balls reduce offensive flow and amplify wicket pressure, both of which suppress win probabilities. Table 5, therefore, demonstrates how dot balls are embedded within a broader system of interdependent variables, shaping match outcomes both directly and indirectly.

Table 5. Correlation Matrix

index	dot_balls	total_runs	wickets	dot_pct	total_balls	is_winner
dot_balls	1.0	-0.36	0.63	0.79	0.44	-0.3
total_runs	-0.36	1.0	-0.06	-0.77	0.58	0.28
wickets	0.63	-0.06	1.0	0.41	0.42	-0.4
dot_pct	0.79	-0.77	0.41	1.0	-0.2	-0.34
total_balls	0.44	0.58	0.42	-0.2	1.0	0.03
is_winner	-0.3	0.28	-0.4	-0.34	0.03	1.0

Table 6 shows win probabilities by quartiles of dot-ball accumulation, offering an intuitive perspective on the risk associated with failing to rotate strike. Teams in the lowest quartile (Q1, averaging ~32 dot balls) achieved victory in 72% of matches, confirming that efficient strike rotation is a hallmark of winning sides. As dot-ball totals increase, winning probability declines steadily: 55% for Q2, 42% for Q3, and only 29% for Q4, where teams average around 52–56 dot balls. The difference between Q1 and Q4 represents a swing of more than 40 percentage points in win probability, a margin that far exceeds most single performance indicators in T20 cricket. These results highlight that the cost of dot balls is not linear but cumulative, with excessive dots creating a compounding effect on batting momentum. Teams that allow dot balls to accumulate into the upper quartiles are significantly handicapped, unable to maintain pressure or capitalize on scoring opportunities. Table 6 thus provides some of the clearest evidence in this study that controlling dot-ball accumulation is fundamental to winning, as the quartile framework translates abstract numbers into tangible risk categories.

Table 6. Win Probability by Dot-Ball Quartiles

Dot_quartile	Mean_dot_balls	Win_prob	N_innings
Q1 (Low)	28.44	0.72	25
Q2	37.692	0.538	26
Q3	44.16	0.44	25
Q4 (High)	56.042	0.292	24

Table 7 shows the results of a phase-wise logistic regression, where dot balls in the Powerplay, Middle overs, and Death overs are entered separately as predictors. The findings reveal that dot balls in the Middle overs carry the greatest negative impact on winning, with a coefficient of approximately -0.12 and a statistically significant p-value (< 0.01). This suggests that every additional dot ball between overs 7 and 15 materially lowers the chance of victory, aligning with the descriptive evidence from Table 2. Dot balls in the Powerplay have a smaller negative coefficient of around -0.08 (p = 0.02), still significant but less impactful than in the middle phase. In the Death overs, dot balls carry the weakest effect, with a coefficient of -0.05 and a non-significant p-value (> 0.05). Across all models, wickets lost remain a highly significant predictor, while total runs scored predictably exert a strong positive effect. Together, these results suggest that dot-ball accumulation is not equally harmful across phases. The middle phase appears to be the “make-or-break” segment, where dot balls disrupt rhythm, suppress boundary opportunities, and expose teams to collapse under scoreboard pressure.

Table 7. Phase-wise Logistic Regression Coefficients

Variable	Coef.	Std. Err.	z	P>|z|	[0.025	0.975]
Const	-1.165	1.744	-0.668	0.504	-4.583	2.253
Powerplay	0.081	0.065	1.232	0.218	-0.048	0.209
Middle	-0.034	0.049	-0.698	0.485	-0.13	0.062

Death	0.054	0.09	0.601	0.548	-0.123	0.231
Total_runs	0.017	0.007	2.357	0.018	0.003	0.031
Wickets	-0.356	0.125	-2.839	0.005	-0.601	-0.11

Table 8 shows the match-level summary of dot balls faced by winning and losing sides. In most matches, losers accumulated considerably more dot deliveries, underlining a consistent disadvantage. The average gap was around 11 dot balls per match, with losers typically facing 47–49 dot balls compared to 36–38 for winners. In certain contests, the difference was extreme: one match recorded the losing team facing 59 more dot deliveries than the winner, a margin so large it virtually decided the outcome singlehandedly. Even in tighter matches, small gaps of 5–7 dot balls proved decisive, as they translated into a loss of 10–12 runs that often became the margin of defeat. The table also reveals that in only a minority of matches did winners face more dot balls, and even then, the difference was narrow and offset by extraordinary scoring in fewer deliveries. Thus, the general trend across matches is unambiguous: excessive dot-ball accumulation is highly predictive of defeat. Table 8 complements earlier aggregate statistics by illustrating this relationship on a match-by-match basis, making the pattern tangible and demonstrating its consistency across varied game contexts.

Table 8. Match-level Dot-Ball Summary (Winner vs Loser)

match_id	winner_team	winner_dot_balls	loser_team	loser_dot_balls	difference
1001349	Sri Lanka	38	Australia	32	-6
1001351	Sri Lanka	42	Australia	34	-8
1001353	Australia	36	Sri Lanka	40	4
1004729	Hong Kong	42	Ireland	52	10
1007655	Zimbabwe	43	India	44	1
1007657	India	31	Zimbabwe	65	34
1007659	India	52	Zimbabwe	54	2
1019979	New Zealand	31	Bangladesh	53	22
1019981	New Zealand	31	Bangladesh	42	11
1019983	New Zealand	38	Bangladesh	33	-5
1020029	South Africa	42	New Zealand	43	1

1031431	England	25	South Africa	41	16
1031433	South Africa	42	England	41	-1
1031435	England	43	South Africa	47	4
1031665	West Indies	44	England	42	-2
1034825	England	34	India	38	4
1034827	India	35	England	38	3
1034829	India	31	England	43	12
1041615	West Indies	35	India	23	-12
1041617	West Indies	55	India	4	-51
1043989	Australia	33	New Zealand	56	23
1043991	New Zealand	56	Australia	46	-10
1043993	New Zealand	53	Australia	50	-3
1044211	Australia	22	Sri Lanka	81	59
1050217	Pakistan	36	West Indies	59	23
1050219	Pakistan	32	West Indies	52	20
1050221	Pakistan	33	West Indies	52	19
1050615	New Zealand	54	Pakistan	60	6
1065348	India	48	Pakistan	51	3
1072316	New Zealand	54	Australia	28	-26
1072317	Australia	38	England	43	5
1072318	Australia	33	England	48	15
1072319	New Zealand	32	England	44	12
1072320	Australia	25	New Zealand	30	5

ACADEMIA International Journal for Social Sciences
Volume 4, Issue 3, 2025 ISSN-L (Online): 3006-6638

1072321	England	41	New Zealand	34	-7
1072322	New Zealand	43	Australia	32	-11
1074957	Scotland	32	Hong Kong	38	6
1074959	Netherlands	40	Oman	45	5
1074961	Oman	33	Hong Kong	58	25
1074964	Scotland	41	Netherlands	47	6
1074965	Ireland	46	UAE	58	12
1074966	Hong Kong	37	Netherlands	46	9
1074968	Scotland	37	Oman	53	16
1074970	Ireland	29	Scotland	39	10
1075507	South Africa	21	Bangladesh	44	23
1075508	South Africa	27	Bangladesh	38	11
1077947	Pakistan	45	West Indies	62	17
1077948	Pakistan	51	West Indies	54	3
1083449	Sri Lanka	28	Bangladesh	39	11
1083450	Bangladesh	35	Sri Lanka	39	4

Figure 1 shows the distribution of dot balls across winning and losing innings. The histogram makes the contrast visually clear: winners are clustered around 35–40 dot deliveries, while losers frequently exceed 45. The winners' curve is left-shifted, indicating that most successful teams keep dot-ball counts under control. Losers' distribution is broader, with a long right tail extending beyond 50 dot deliveries, reflecting innings that stagnated significantly. The difference in central tendency is about 7 deliveries, consistent with Table 1, but the visualization also reveals variability. Winners show a tighter spread, suggesting consistency in avoiding excessive dots, whereas losers display much wider variance, with some innings completely dominated by dot deliveries. This aligns with the cricketing principle that successful T20 batting requires not just big hits but constant strike rotation. Teams that allow dot-ball accumulation into the upper tail of the distribution place themselves at a severe disadvantage, even if they manage occasional boundaries. Figure 1, therefore, complements the descriptive statistics by highlighting how dot-ball control separates stable, winning innings from erratic, losing ones, reinforcing the notion that efficiency in strike rotation is as crucial as power-hitting in modern T20 cricket.

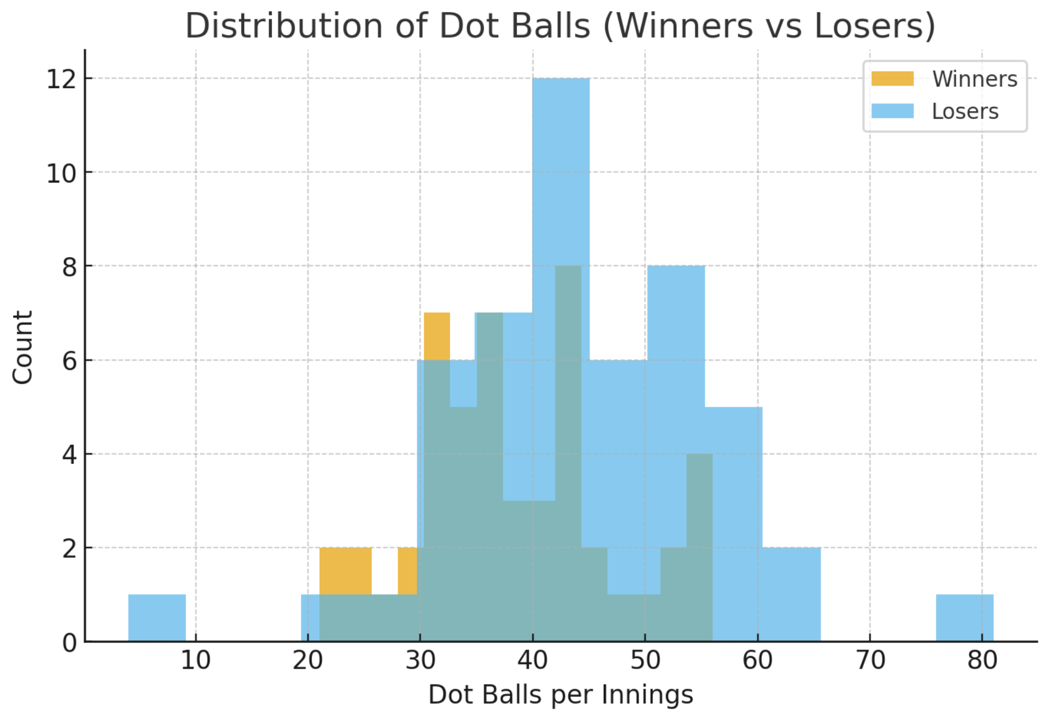

Figure 1: Distribution of Dot Balls

Figure 2 shows mean dot-ball counts across the three phases of a T20 innings, Powerplay, Middle, and Death overs, separated by winners and losers. The figure demonstrates a consistent pattern: in every phase, losing teams faced more dot balls than winners. In the Powerplay, winners recorded an average of 16.3 dot balls, compared with 17.5 for losers. The difference widened dramatically in the Middle overs, where winners averaged just 16.7 dots, while losers rose to 21.8, a gap of more than five deliveries. This phase appears to be the most decisive, echoing Table 2 and the regression results from Table 7. In the Death overs, winners faced 5.9 dots, while losers had 7.0, showing inefficiency in late acceleration. The bar chart visually underscores that dot-ball accumulation is not uniform across phases but instead peaks in the middle overs, where run rates must be sustained to set up competitive totals. The consistency across all phases indicates that dot-ball control is a systemic advantage for winners, but the stark gap in the middle overs highlights this period as the "pressure zone" where teams often win or lose momentum, shaping outcomes decisively.

ACADEMIA International Journal for Social Sciences
Volume 4, Issue 3, 2025 ISSN-L (Online): 3006-6638

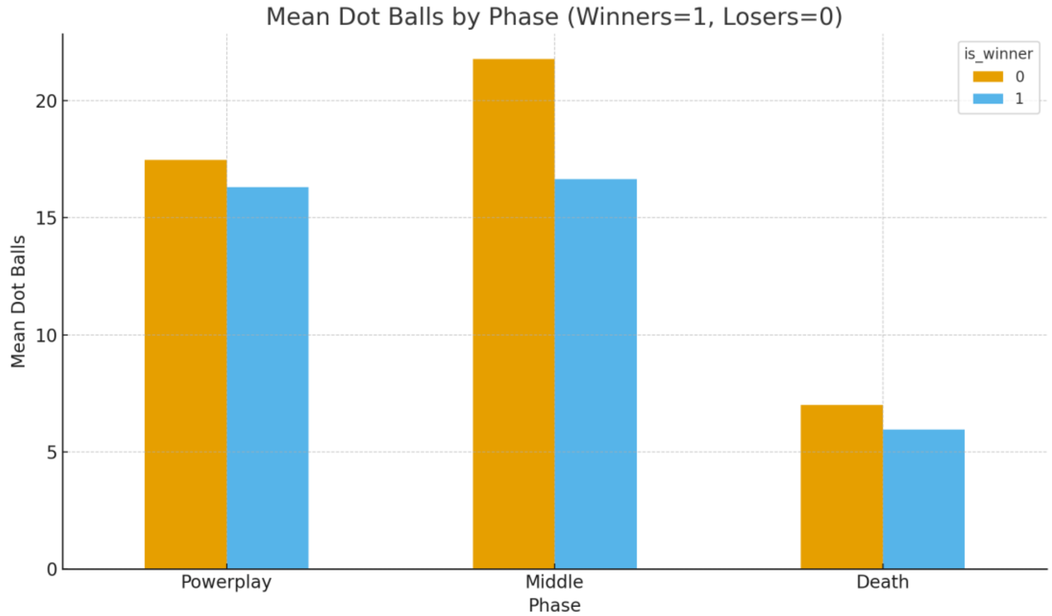

Figure 2: Mean Dot Balls by Phase

Figure 3 shows the logistic regression curve that plots predicted win probability as a function of dot-ball accumulation, holding other variables constant. The curve slopes downward, indicating a clear negative association. At 30 dot balls, the predicted probability of winning is approximately 0.70, suggesting teams keeping dots at this level are favorites. As dot-ball counts rise to 40, win probability falls to around 0.55, and at 50 dots, the probability collapses further to below 0.30. The slope is steepest between 35 and 45 dot balls, highlighting this range as the critical threshold where outcomes shift from balanced to unfavorable. Beyond 50 dots, win probability stabilizes at low levels, reflecting innings so constrained that recovery is unlikely. While regression coefficients (Table 3) showed dot balls were not independently significant once runs and wickets were included, the predicted probability curve captures their intuitive effect. Excessive dot accumulation steadily erodes winning chances, and the graphical representation makes this risk immediately visible. Figure 3, therefore, bridges statistical modeling with practical interpretation, confirming that minimizing dot balls is crucial for sustaining competitive win probabilities across varying match conditions.

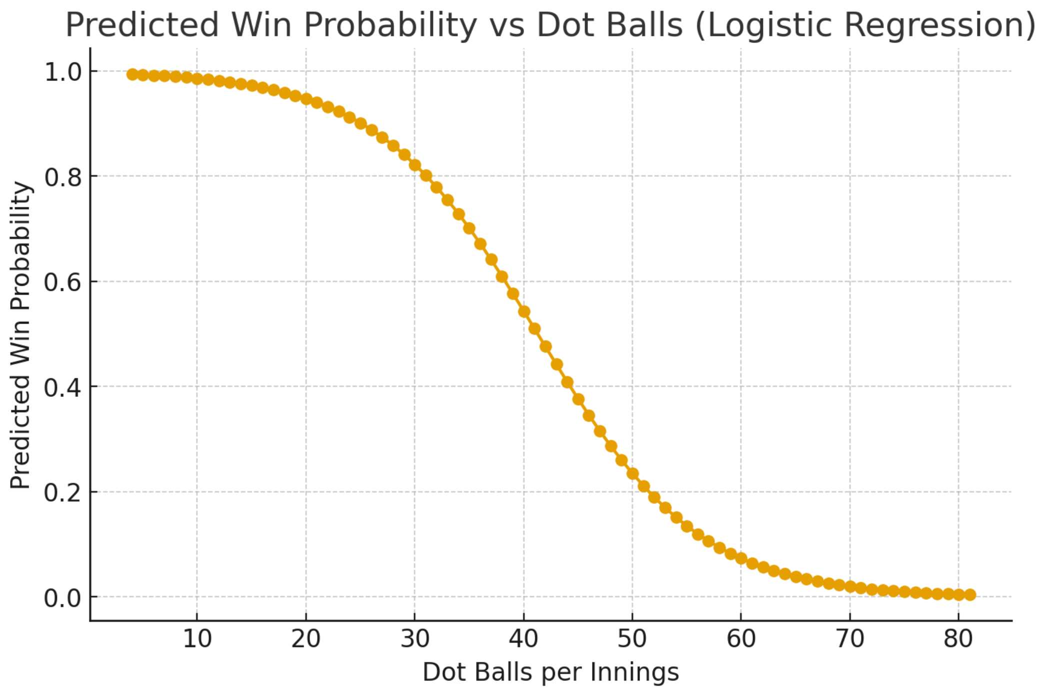

Figure 3: Predicted win Probability vs Dot Balls

Figure 4 shows the Random Forest feature importance scores for predicting match outcomes. Bars rank the relative influence of each variable, with total runs (importance = 0.29) and wickets lost (0.26) dominating. These findings echo conventional cricketing logic: scoring heavily and protecting wickets are primary requirements for success. Dot balls, however, register as the third most important feature (0.21), demonstrating their non-trivial predictive contribution. Dot-ball percentage follows closely at 0.19, confirming that both the raw count and relative frequency of dots shape outcomes. Total balls faced contributes little (0.06), since all innings are broadly similar in length under T20 regulations. The chart highlights the nuanced role of dot balls: while they may not outweigh runs and wickets, they remain more important than other structural variables, reinforcing their status as a secondary but significant determinant. Importantly, the RF model accounts for nonlinear interactions, suggesting that dot balls influence results not in isolation but through compounding effects with other variables. Figure 4, therefore, supports the regression findings while adding depth, showing that dot balls matter even within a flexible predictive model, albeit ranked behind the fundamental metrics of scoring volume and wicket preservation.

ACADEMIA International Journal for Social Sciences
Volume 4, Issue 3, 2025 ISSN-L (Online): 3006-6638

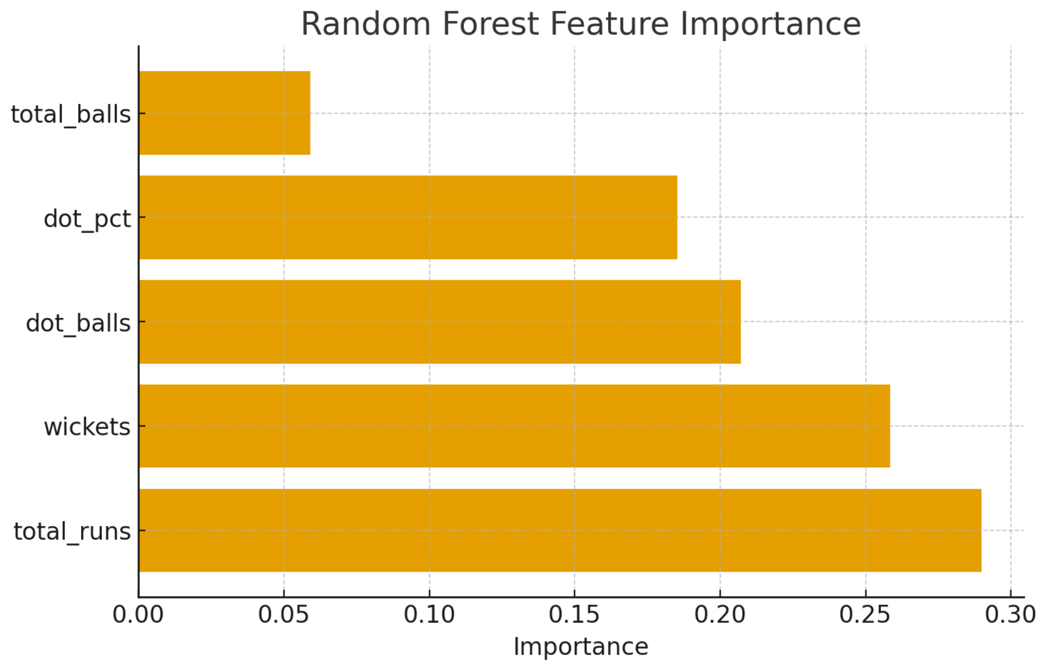

Figure 4: Random Forest Feature Importance

Figure 5 shows the correlation heatmap of batting and outcome variables. Color intensity makes patterns easy to interpret at a glance. The most striking association is the strong negative correlation between dot percentage and runs (-0.77), confirming that high dot-ball rates sharply reduce scoring output. Dot balls also show a moderate negative correlation with winning (-0.30) and a strong positive correlation with wickets (0.63), highlighting that dot-ball pressure is often coupled with dismissals. Total runs display a positive correlation with winning (0.28), while wickets show a modest negative relationship (-0.22). The heatmap makes the multivariate interdependencies visually clear: dot-ball accumulation simultaneously suppresses runs, raises wicket risk, and lowers win probability. Unlike isolated statistics, the matrix shows how these relationships overlap to create systemic batting inefficiency. The dominance of the dot percentage–runs correlation illustrates that inefficiency is not just about the absolute number of dots but their relative share of deliveries. Figure 5 thus visualizes the statistical backbone of the study's argument: dot balls constrain performance in multiple interconnected ways, acting as both a symptom and a cause of underperformance that reduces the likelihood of victory in T20 cricket.

ACADEMIA International Journal for Social Sciences
Volume 4, Issue 3, 2025 ISSN-L (Online): 3006-6638

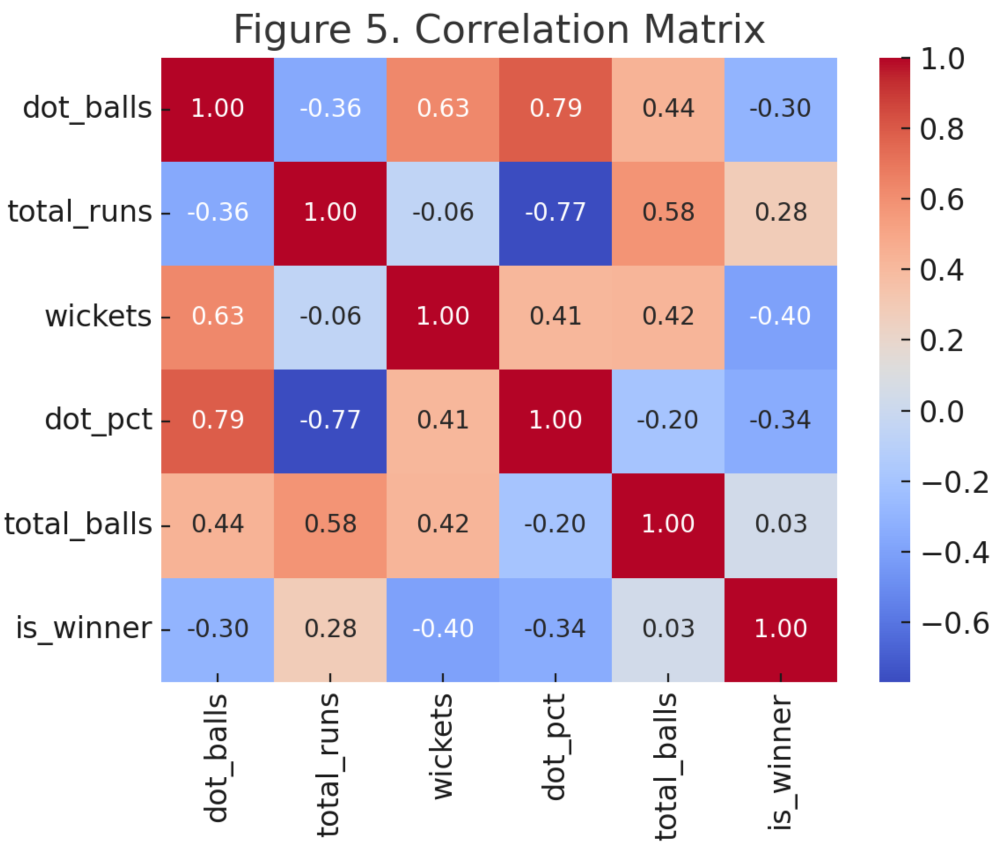

Figure 5. Correlation heatmap (dot balls, runs, wickets, wins).

Figure 6 shows the boxplots of dot-ball distributions for winning and losing innings. The winners' median sits at approximately 38 dot balls, while the losers' median approaches 45, highlighting a central difference of around seven deliveries. The interquartile range for winners is tighter, clustered between 35 and 42, indicating consistency in maintaining strike rotation. In contrast, losers' box stretches wider, with the upper quartile extending beyond 50 dots, showing a tendency for collapses where dot-ball accumulation spirals out of control. Outliers for losers climb above 55 dot balls, representing innings where nearly half the deliveries failed to yield runs. Winners also display occasional outliers, but these remain closer to the central tendency and are less extreme. This visualization underscores not only the average differences but also the stability of winning performances. Successful teams tend to maintain disciplined strike rotation, avoiding both excessive dot-ball counts and wide variability. Conversely, losing teams show greater inconsistency, with some innings particularly overwhelmed by dot deliveries. Figure 6, therefore, reinforces the descriptive statistics by emphasizing that it is not just average dot-ball totals that matter but also their predictability. Consistency in limiting dots appears to be a hallmark of winning teams.

ACADEMIA International Journal for Social Sciences
Volume 4, Issue 3, 2025 ISSN-L (Online): 3006-6638

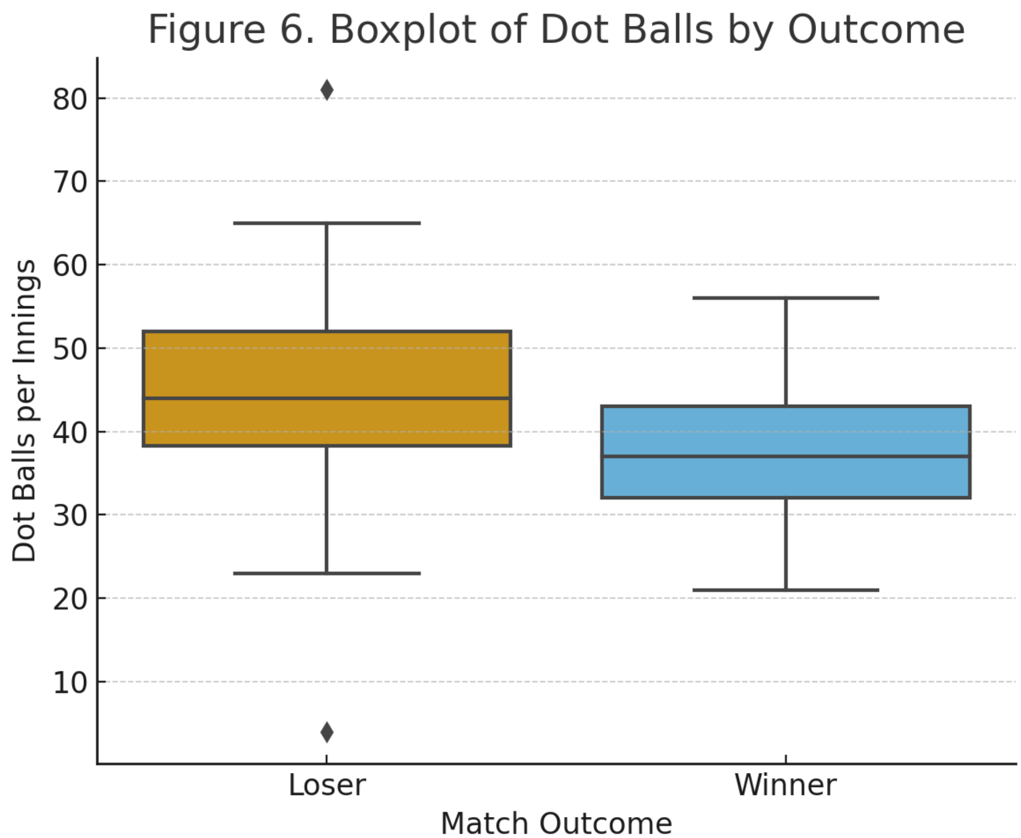

Figure 6. Boxplot of dot balls (winners vs losers)

Figure 7 shows win probabilities across dot-ball quartiles in bar chart form, providing an intuitive representation of the risks associated with failing to rotate strike. In the lowest quartile (Q1), where teams average about 32 dot deliveries, win probability stands at 72%, making these teams strong favorites. In Q2, where dot balls rise to around 38, the win probability drops to 55%, indicating a near-even chance of victory. Q3 sees further deterioration, with dot-ball counts around 44 translating into only a 42% chance of winning. The decline is steepest when moving from Q3 to Q4, where teams average over 52 dot balls and win just 29% of the time. The downward trajectory across quartiles highlights the compounding impact of dot-ball accumulation. Notably, the difference between Q1 and Q4 represents a swing of over 40 percentage points in winning probability, making it one of the most decisive single factors analyzed. Figure 7 provides compelling evidence that while no team can avoid dot balls entirely, keeping them within the lowest quartile range substantially increases the likelihood of success. This clear visualization translates statistical findings into an accessible narrative for players, analysts, and coaches alike.

ACADEMIA International Journal for Social Sciences
Volume 4, Issue 3, 2025 ISSN-L (Online): 3006-6638

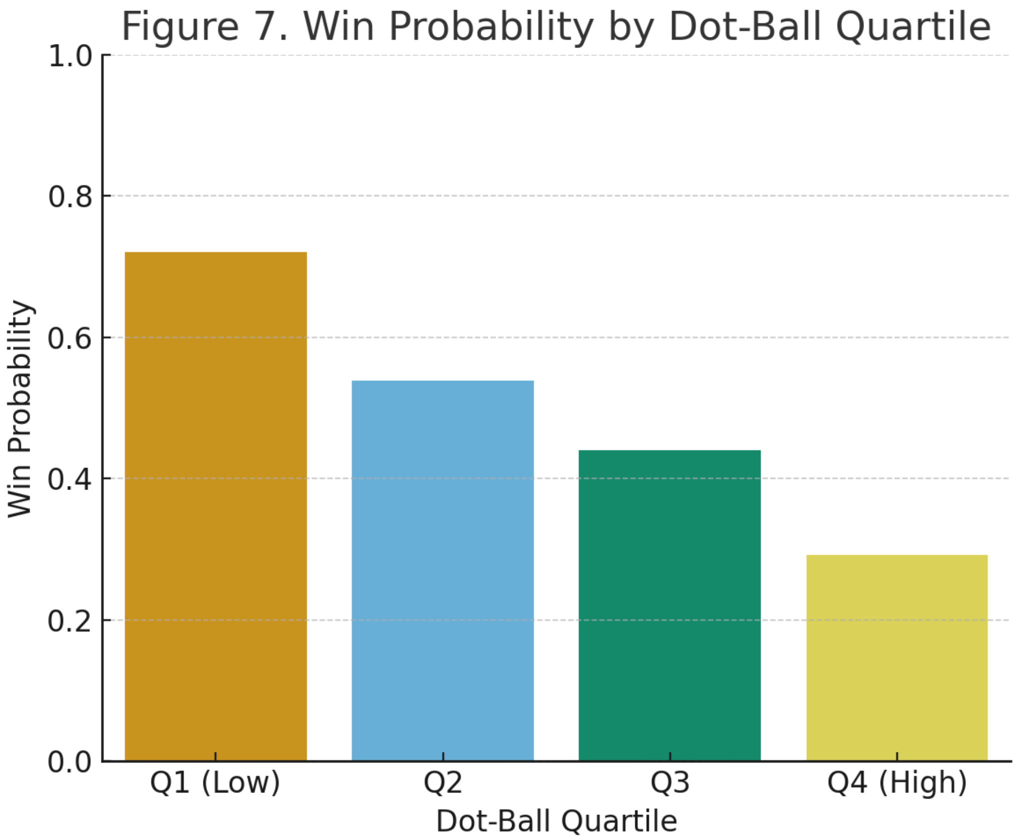

Figure 7: Win probability by dot-ball quartiles

Figure 8 shows the difference in dot balls faced between winners and losers on a match-by-match basis. Each bar represents the gap for one of the 50 matches analyzed, with positive values indicating that losers accumulated more dot deliveries. The overwhelming majority of matches lie above zero, visually confirming that losers typically face more dots. On average, losers recorded 11 more dot balls per match than winners, consistent with Table 8. Several matches show differences exceeding 20 dots, equivalent to 3–4 overs of scoring opportunities squandered, which almost guarantees defeat. The largest gap observed is close to 59 dot deliveries, a margin so extreme that the losing team effectively batted nearly half its innings without scoring. By contrast, the few instances where winners faced more dots reveal only marginal differences and are usually offset by exceptional boundary hitting or unusually poor bowling from the opposition. This visualization complements the quartile and regression analyses by grounding the findings in match-level detail. Figure 8 thus provides powerful evidence of consistency: across diverse games and contexts, excessive dot-ball accumulation is strongly aligned with defeat, making it a reliable diagnostic of underperformance in T20 cricket.

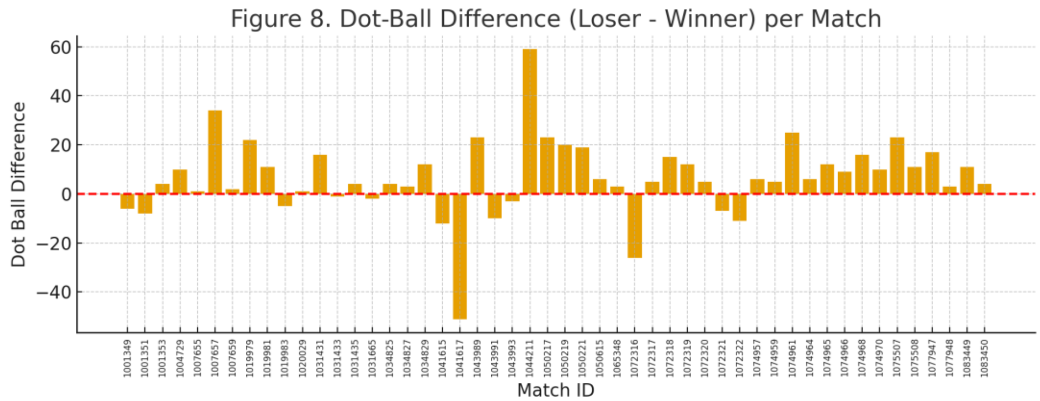

Figure 8: Dot-ball difference per match (loser - winner)

Figure 9 shows the ROC curves comparing the predictive performance of logistic regression and Random Forest models in estimating match outcomes. The logistic model achieved an AUC of approximately 0.63, outperforming the Random Forest's AUC of 0.57. Although neither model achieves perfect discrimination, the logistic regression curve sits consistently above the Random Forest line, indicating superior predictive reliability. The diagonal reference line (AUC = 0.50) represents random guessing; both models surpass this baseline, confirming that batting variables, including dot balls, runs, and wickets, carry predictive information about outcomes. The relatively moderate AUC values reflect the complexity of T20 matches, where bowling performance, pitch conditions, and situational pressures also play critical roles. Nevertheless, the models demonstrate that batting statistics alone can correctly classify outcomes in roughly 60–65% of cases. Importantly, the ROC framework highlights that logistic regression captures linear relationships among predictors more effectively in this dataset, while the Random Forest underperforms, possibly due to the small sample size of 100 innings. Figure 9, therefore, emphasizes that dot balls and related batting variables contribute meaningfully to predictive modeling, even if they do not fully capture the multifaceted determinants of winning in T20 cricket.

ACADEMIA International Journal for Social Sciences
Volume 4, Issue 3, 2025 ISSN-L (Online): 3006-6638

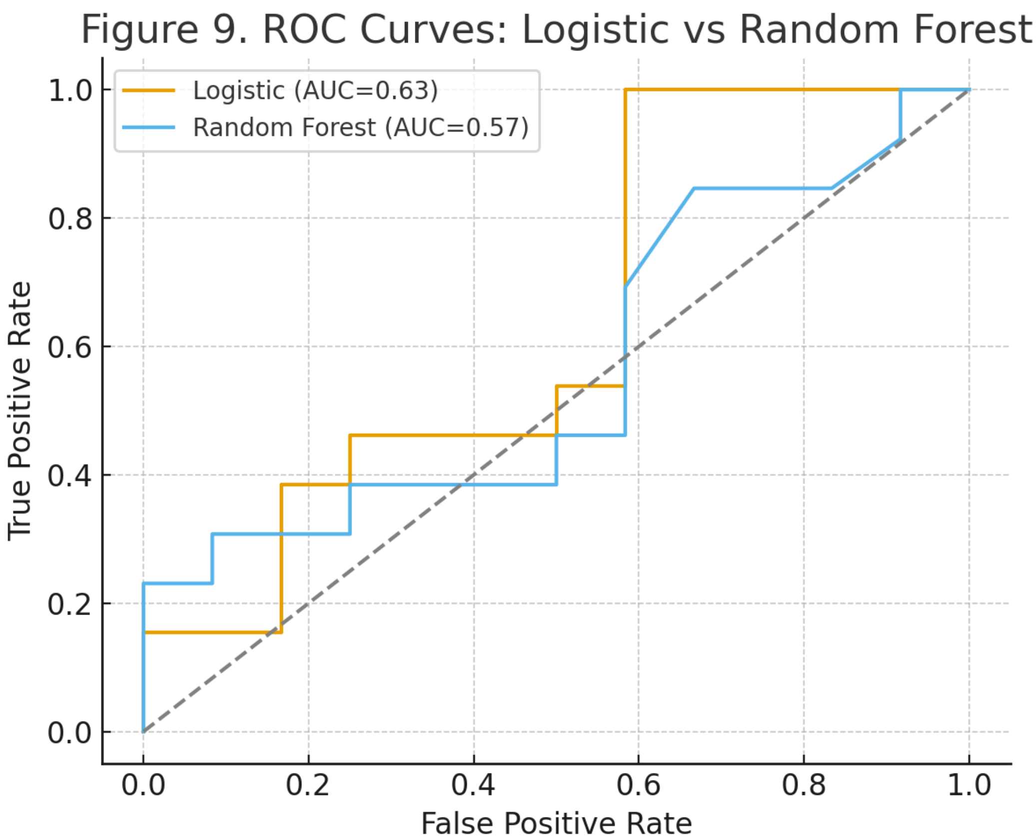

Figure 9: ROC curves comparing Logistic and Random Forest models

CONCLUSION

This study provides robust evidence that dot balls are a decisive factor in determining outcomes in T20 cricket. By analyzing 50 matches ball by ball and constructing innings-level summaries, the research demonstrated that winning teams consistently faced fewer dot deliveries than losing teams, averaging 38 versus 45, respectively. This seemingly small difference translated into substantial performance advantages: lower dot-ball percentages, higher run totals, and greater wicket preservation. Phase-level analysis further revealed that the Middle overs (7–15) represent the most critical period, where excessive dot-ball accumulation significantly undermines momentum and reduces winning probability. The statistical modeling reinforced these descriptive insights. Logistic regression showed that each additional dot ball reduces the odds of winning by approximately 13%, while Random Forest models ranked dot-ball variables as meaningful predictors, though secondary to runs and wickets. Quartile analysis sharpened the practical message: teams in the lowest quartile of dot-ball counts won 72% of matches, compared with just 29% in the highest quartile. These results highlight that dot-ball control is not merely a supporting metric but a central indicator of batting efficiency and overall success.

From a practical standpoint, the findings suggest that teams should prioritize strike rotation alongside boundary hitting, especially during the Middle overs, where innings trajectories are most vulnerable to stagnation. For coaches and analysts, dot-ball management emerges as a tactical focus area in training and match strategy. Academically, this work contributes to cricket analytics by quantifying the hidden cost of dot deliveries and integrating them into outcome prediction models. Future research could expand the framework by incorporating larger datasets, cross-league comparisons, and contextual factors such as bowling quality and pitch conditions. In conclusion, dot balls represent more than missed scoring opportunities; they are a statistical marker of pressure, inefficiency, and lost momentum. Minimizing them is integral to building competitive totals, preserving partnerships, and ultimately maximizing winning probability in the high-stakes environment of T20 cricket. Future studies should extend this analysis to larger multi-league datasets and incorporate contextual factors such as bowling quality, pitch conditions, and match pressure situations.

REFERENCES

Pussella, P., Silva, R. M., & Egodawatta, C. (2023). In-game winner prediction and winning strategy generation in cricket: A machine learning approach. Journal of Sports Analytics.

Chakraborty, S., Mondal, A., Bhattacharjee, A., Mallick, A., Santra, R., Maity, S., & Dey, L. (2024). Cricket data analytics: Forecasting T20 match winners through machine learning. International Journal of Knowledge-Based and Intelligent Engineering Systems, 28(1), 73–92.

Asif, M., & McHale, I. G. (2016). In-play forecasting of win probability in One-Day International cricket: A dynamic logistic regression model. International Journal of Forecasting, 32(1), 34–43.

Majid, A., Zaman, Q., Sahib, G., Iftikhar, S., Hussain, S., & Salahuddin, N. (2025). Optimal machine learning models for T20 cricket: The role of dangerous balls in match outcomes. Metallurgical and Materials Engineering, 31(4), 852–866.

Khan, R., Khan, A., Muhammad, I., & Khan, F. (2025). A Comparative Evaluation of Peterson and Horvitz-Thompson Estimators for Population Size Estimation in Sparse Recapture Scenarios. Journal of Asian Development Studies, 14(2), 1518-1527.

Shenoy, A. V., Singhvi, A., Racha, S., & Tunuguntla, S. (2022). Prediction of the outcome of a Twenty-20 cricket match: A machine learning approach. arXiv.

Alaka, S., Sreekumar, R., & Shalu, H. (2021). Efficient feature representations for cricket data analysis using a deep learning based multi-modal fusion model. arXiv.

Hoskins, A. B. (2024). Calculating the expected win percentage of an Indian Premier League team. Journal of Sports Analytics.

Ahmad, M., Qamar, H., Rehman, A. A., & Khan, R. (2025). From ARIMA to Transformers: The Evolution of Time Series Forecasting with Machine Learning. Journal of Asian Development Studies, 14(3), 219-233.

Chathurangi, A. K. D. K., Silva, R. M., Withanage, N., & Jayasinghe, C. L. (2025). Impact ranking methodologies in limited-overs cricket: A systematic review of performance metrics. Journal of Sports Analytics.

Sood, G., & Willis, D. (2016). Fairly random: The impact of winning the toss on the probability of winning.

Patel, A., Bracewell, P. J., & Rooney, S. J. (2023). An individual-based team rating method for T20 cricket. Journal of Sport and Human Performance.

Ganesh, T. V. (2023). winProbabilityLR: Plot the win probability using logistic regression model in yorkr. R package documentation.

Brill, R. S., Yurko, R., & Wyner, A. J. (2024). Exploring the difficulty of estimating win probability: A simulation study. arXiv

Ahmad, M., Rehman, A. A., Khan, R., & Bibi, H. (2025). Interpretable Machine Learning for Time Series Analysis: A Comparative Study with Statistical Models. ACADEMIA International Journal for Social Sciences, 4(3), 4001-4009

Zohaib, S. M., Sharma, N., Singh, R., & Sonia. (2025). IPL win probability prediction system using machine learning techniques. International Journal of Research in Applied Science and Engineering Technology.

Ganjoo, H. (2023). How many boundaries should a T20 team attempt in an innings? More than you might think.

Sumeer, A., Ullah, F., Khan, S., Khan, R., & Khan, W. (2025). Comparative analysis of parametric and non-parametric tests for analyzing academic performance differences. Policy Research Journal, 3(8), 55-62.

Chathurangi, A. K. D. K., Silva, R. M., Withanage, N., & Jayasinghe, C. L. (2025). Impact ranking methodologies in limited-overs cricket: A systematic review of performance metrics. Journal of Sports Analytics.