Journal of Technology and Value Addition, Volume 7 (2), July, 2025: (54-69)

Research Article

A Data-Driven Approach to Quantifying the Effect of Extra Deliveries (Wides and No-Balls) in T20 Cricket

P.P.G. Dinesh Asanka*

dasanka@kln.ac.lk

Department of Industrial Management, University of Kelaniya, Kelaniya, Sri Lanka

Abstract

Cricket experts often claim that extra deliveries (wides or no-balls) disproportionately benefit the batting team. This study challenges that assumption by analyzing the relationship between extra deliveries and their outcomes in T20 cricket using a data-driven approach. Employing nine classification algorithms and six feature selection techniques, it was found that there was no significant correlation between extra delivery outcomes and other match parameters. While Naive Bayes showed slightly better predictive accuracy, feature selection methods (especially Chi-Squared) identified match stage, batsman strike rate, and runs conceded by the bowler as potential factors. When classifications were executed the selected features, the comparative analysis of machine learning algorithms reveals that Random Forest achieves the highest AUC (60) and MCC (5), demonstrating robust performance across evaluation metrics. Notably, Logistic Regression attains the best classification accuracy (69) but suffers from low precision (47), indicating a trade-off between true positives and false positives. However, classification models did not consistently support these relationships. Research findings suggest that extra delivery outcomes are largely unpredictable, resembling random events rather than predictable results influenced by specific match factors. This research provides a more nuanced understanding of the impact of extra deliveries in T20 cricket, emphasizing the need for caution when interpreting expert opinions.

Keywords: *Azure Machine Learning, Classification, Cricket, Feature Selection, Orange Data Mining*

1. Introduction

Cricket, with an estimated 2.5 billion fans worldwide, stands as the second most popular sport after soccer (Top-10 Most Popular Sports in the World 2023, 2023), captivating audiences primarily across Asia. Distinguishing itself from other sports, cricket boasts three unique formats: Test matches, 50-over matches, and Twenty20 (T20) matches (International Cricket Council, n.d.). The advent of franchise cricket has propelled T20 cricket to the forefront of the sport.

Under the laws of cricket, a bowler typically delivers six legitimate deliveries in an over. However, in cases of a "no-ball" (overstepping the crease) or a "wide" (bowling out of reach for the batsman), an additional delivery must be bowled. Cricket experts commonly believe that these extra deliveries disproportionately favor the batting team. While it's evident that an extra delivery contributes to the batting team's score, this research aims to investigate whether the advantage is as substantial as experts claim.

This study will analyze the impact of extra deliveries on considering factors such as the format of the game, the stage of the match, and the team's overall performance. The findings will shed light on the validity of the prevailing belief and contribute to a deeper understanding of the intricacies of cricket strategy.

With the increasing commercialization of cricket, the sport, like many others, has embraced predictive and prescriptive analytics over traditional descriptive and diagnostic approaches. This study harnesses machine learning techniques, employing Azure ML Classic (Microsoft, n.d.) and Orange Data Mining (Orange Data Mining, n.d.), to delve into the impact of extra deliveries. The research analyzes various dimensions, including gender, international versus franchise cricket, bowler and batsman parameters, and match status, to provide a comprehensive understanding of the phenomenon. This multi-faceted analysis will enhance the comprehension of how extra deliveries influence match outcomes across different contexts.

2. Literature Review

Analyzing the impact of a wide or no-ball delivery in T20 matches involves understanding how these events influence the match outcome using classification and feature selection techniques. Various studies have demonstrated the effectiveness of machine learning models in predicting cricket match outcomes by leveraging different feature selection methods. For instance, Recursive Feature Elimination and XG Boost have been applied to predict match outcomes with high accuracy, indicating the importance of selecting relevant features for model performance (Shakil *et al.,* 2020). Similarly, the LASSO method has been used to identify crucial features in predicting the winner of T20 matches, showing that feature selection significantly enhances prediction accuracy (Pussella *et al.,* 2023). In the context of cyberstalking detection, the TF-IDF + Chi-Square Test approach has proven effective in improving model performance by selecting relevant features, which can be analogously applied to cricket match data to filter out non-influential features (Gautam & Bansal, 2022). Additionally, the use of clustering techniques combined with feature selection methods like Information Gain has shown to improve classification performance in software defect prediction, suggesting that similar approaches could be beneficial in analyzing cricket match data (Usman-Hamza *et al.,* 2019). Studies on the Bangladesh Premier League have highlighted that ensemble classifiers like Gradient Boosting outperform other models when considering all features, emphasizing the need for robust feature selection to enhance prediction accuracy (Pramanik *et al.,* 2022). Furthermore, the application of optimization algorithms like Particle Swarm Optimization (PSO) for feature selection has yielded convincing results in various datasets, indicating its potential utility in cricket match analysis (Ajibade *et al.,* 2021). By integrating these methodologies, one can effectively analyze the impact of wide or no-ball deliveries on the outcome of T20 matches, providing valuable insights for teams and analysts to devise better strategies and improve decision-making processes.

The increasing emphasis on tactical strategies in sports has led to the widespread use of statistical analysis in cricket to enhance decision-making across various domains. Cricket analytics encompass diverse areas, including player performance evaluation, team selection, strategic guidance for coaches and managers, injury prevention and management, analysis of umpiring decisions, score prediction, personalized recommendations, assessment of pitch and ground conditions, and evaluation of the impact of match interruptions (Wickramasinghe, 2022).

The popularity of fantasy cricket has spurred significant research aimed at improving the accuracy of predicting winning teams. A study by (Kumar *et al.,* 2022) employed various classifier techniques, including k-Nearest Neighbors (KNN), Naive Bayes, Support Vector Machines (SVM), Decision Trees, and Random Forests, along with extensive data cleaning and engineering. However, the study's evaluation focused solely on accuracy, neglecting other important classification metrics like recall, precision, and F1 score.

A dedicated book chapter (Mathew *et al.*, 2023) on cricket analytics emphasized the importance of robust data engineering as a foundation for successful machine learning applications. This chapter examined various match-related data points, including toss outcomes, match results, winning margins, and venues, employing linear and logistic regression models to forecast individual player scores.

Furthermore, researchers have utilized visual representations such as contour plots (Bhattacharjee & Saikia, 2023) to analyze bowlers' performance metrics, including bowling averages, strike rates, and economy rates for international cricketers in One Day Internationals (ODIs). A broad range of statistical (Saraswat *et al.,* 2018) and machine learning (Rahman *et al.,* 2018) methods have been applied to investigate numerous aspects of cricket, enhancing the comprehension of the sport's intricacies.

Research conducted on male cricketers in England and Wales (Kelly *et al.,* 2022) highlighted the significance of diversity factors such as gender and age group. In addition, a separate study (Anuraj *et al.,* 2023) identified winning the toss and team rankings as key parameters for predicting outcomes in T20 international matches.

Beyond the parameters identified above, research in cricket extends to various other areas, including the performance of out-of-form players (Silva *et al.*, 2022), the impact of power play strategies (Silva, n.d.), and the survival abilities of batsmen (Kottearachchi *et al.*, 2022).

Regarding the impact of extra deliveries in cricket, the author of this paper has previously conducted two studies. One study employed data warehousing and data mining techniques (Asanka, 2014) to assess this impact. However, that study was conducted roughly ten years ago, and T20 cricket has evolved considerably since then, largely due to the proliferation of global franchise leagues and heightened competition. Additionally, the previous study focused solely on men's cricket, while women's cricket has also seen significant advancements in the past decade. The author also performed a descriptive analysis (Asanka, 2023) on the same dataset. This current research expands upon that descriptive analysis with a more comprehensive dataset.

3. Methodology

For this study, data was collected on T20 matches, encompassing international, franchise, men's, and women's matches. Test and fifty-over matches were excluded due to their increased complexity. The reputable website ESPNCricinfo (ESPNCricinfo n.d.) served as the primary data source. Semi-automated web scraping techniques were used to extract data from commentary of matches. Preprocessing tasks were executed to clean data. To account for varying playing conditions in franchise cricket, only matches played in 2023 were included. Furthermore, due to differing rules across franchises, such as point schemes in the South Africa league (SA20) and the "Power Surge" in Australia's Big Bash League (BBL), specific franchises and overs were omitted. Similarly, overs affected by the DLS (Duckworth-Lewis-Stern) method (Duckworth & Luwis, 2009) were excluded to ensure a fair comparison. Additionally, incidents with multiple extra deliveries in an over, as well as extras in half-completed overs finished by multiple bowlers, were disregarded.

The outcome of the last delivery in a T20 cricket match can be influenced by a multitude of factors, including the performance statistics of the players, the situational context of the match (Saikia & Bhattacharjee, 2010). Environmental factors like weather conditions and the time of the match (day or day-night) can also impact the last delivery, as they affect visibility and pitch conditions (Dasgupta, 2016). The psychological pressure on both the batsman and the bowler during the final moments of the game can significantly influence their performance (Callaghan, 2018). All these factors combined create a complex and dynamic environment that determines the outcome of the last delivery in a T20 cricket match. For the dataset, various parameters were considered that were referred to in the research papers and expert opinion. Such as batsman score, deliveries faced, boundaries hit, balls delivered by the bowler, runs conceded, wickets taken, extra ball number, runs in the previous and subsequent overs, runs in the current over, extra type (wide/no-ball), change of batsman during a free-hit, innings number, over range (power play, middle, final), partnership details (runs, deliveries, presence of a wicket in the same over), gender, and match type (international/franchise). Numerical data was categorized using PowerBI (Microsoft Power BI, 2024) to facilitate classification in Orange Data Mining (Orange Data Mining, n.d.), while the original numerical data will be utilized for future analysis in Smart PLS. The original and transformed datasets, along with PowerBI and Orange Data Mining files, are available on GitHub (Asanka, n.d.) for public access. Additionally, the AzureML experiment (Microsoft, n.d.) has been published for public use to enable further exploration and enhancement of this research.

4. Dataset

This research was done with 269 matches and 1,581 extra delivery instances. The distribution of matches is shown in Table 1.

Table 1: Distribution of Matches

Gender	Match Type		
	International	***Franchise***	***Total***
Male	40	161	201
Female	36	32	68
Total	76	193	269

Table 1 indicates the popularization of franchise Cricket over international Cricket in T20 Cricket.

Table 2: Distribution of Extra Delivery

Gender	Match Type		
	International	***Franchise***	***Total***
Male	225	1,017	1,242
Female	185	154	339
Total	410	1,171	1,581

Out of the 1,581 extra deliveries, wide deliveries accounted for 90%, while only 10% were for no-ball deliveries. Since there are six deliveries per over, it is important to understand delivery distribution which is shown in Figure. 1.

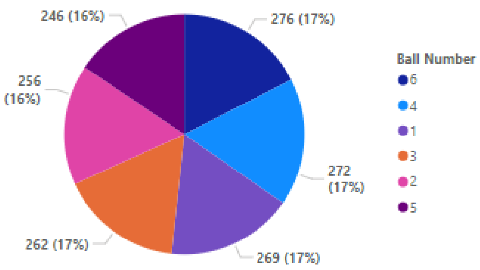

Figure 1: Distribution of deliveries of extra deliveries

Deliveries are generally evenly distributed across all balls in an over. However, there is a slight increase in the frequency of wide deliveries for the first and sixth balls. This could potentially be attributed to a momentary lapse in concentration by the bowlers.

The outcome of each delivery was categorized from the batting team's perspective into five levels of impact: High, Medium, Less, Negative, and No Impact, as illustrated in Figure 2.

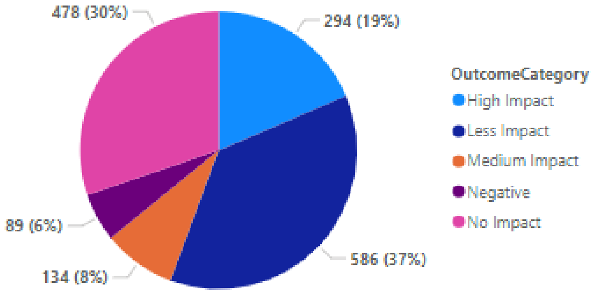

Figure 2: Impact of extra deliveries

"No Impact" indicates that no runs were scored on the extra delivery, while "Negative Impact" signifies that a wicket was taken. As shown in Figure 2, only 20% of extra deliveries resulted in a major impact, while 30% had no impact. Additionally, a 6% negative impact rate is evident in the figure.

Figure 3 illustrates the distribution of extra deliveries across different stages of the match.

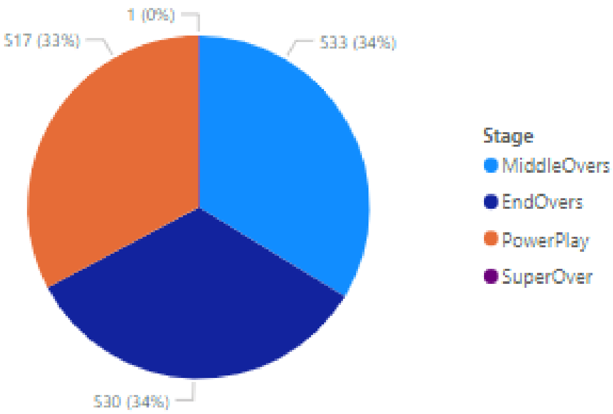

Figure 3: Occurrence of the Extra delivery with respect to the match stage

Given that the match stage was found to be a significant factor, it was observed that the collected dataset had an even distribution of extra deliveries across Power Play, Middle Overs, and End Overs. There was a single instance of an extra delivery in a Super Over between Sri Lanka and New Zealand, which was deemed negligible.

To enhance clarity and interpretability, most of the collected numerical data were transformed into categorical values. With the exception of Innings, IsFreeHit, Gender, Match Type, and Bowler Over Number, the remaining parameters were converted into categorical variables based on their domain or collected values, as detailed in Table 3.

In Table 3, the class variable is the Outcome, while all the other variables are considered to be the independent variables.

Table 3: Dependent Variable for the Outcome of the Last Delivery

Selected Parameters to Determine the outcome of the Extra Delivery Parameter	Value(s)	Value Range
Innings	1,2	
IsFreeHit	Yes/No	
Gender	Male / Female	
Match Type	International / Franchise	
Match Stage	Power Play	1 - 6
	Middle Overs	7-14
	End Overs	15-20
Bowler Over Number	1,2,3,4	
Bowler – Wickets	No Wickets	0
	Low	1-2
	Medium	3-4
	High	>=5
Bowler – Runs Conceded	Low	<= 10
	Medium	11 – 30
	High	>30
Bowler – Economy	Low	<6
	Medium	6.01 – 10
	High	>10
Batsman – Score	Low	<= 20
	Medium	21 – 50
	High	>50
Batsman – Deliveries Faced	Low	<=10
	Medium	11 – 40
	High	>40
Batsman Boundaries	Low	0 – 1
	Medium	2-5
	High	>5
Batsman – Strike Rate / Partnership – Strike Rate	Low	=<100
	Medium	101 -175

	High	>175
	Low	=<40
Partnership Runs	Medium	41-75
	High	>75
	Low	<=30
Partnership – Deliveries	Medium	31-50
	High	>50
Previous Over – Runs / After Over – Runs / Current Over-Runs	Low	=<6
	Medium	6-10
	High	>10
	Frontline	1-3
Partnership Stage	Middle Order	4-7
	Tailend	8-10
	Negative	Wicket
	No Impact	0 runs
Outcome	High Impact	3-6 runs
	Less Impact	1 run
	Medium	2 runs

5. Results

After the descriptive analysis was done in the previous research, this research aims to find the relationship using Machine Learning techniques. In this research, machine learning algorithms were used to decide whether there is a correlation between the delivery outcome and other selected parameters.

Orange data mining was used with nine classification algorithms such as KNN, Random Forest. Naïve Bayes, Ada Boost, SVM, Gradient Boosting, XGBooost, Neural Network, Decision Tree, and Logistic Regression. Cross validation was used as an evaluation technique to overcome overfitting issues. Startified, 5 folders were used in the cross validation evaluation technique. Evaluation parameters for each algorithm are shown in Table 4 In this classification, Area Under the Curve (AUC), Classification Accuracy (CA), F1 score, Precision, Recall, and Mathews Correlation Coefficient (MCC) were compared.

As shown in Table 4, Classification Accuracy, F1, Precision, and Recall parameters are less than 40% while MCC is less than 15%. This indicates that it is impossible to predict the outcome of the delivery when there is an extra delivery. This result confirms the descriptive analytics findings that were carried out in previous research.

Table 4: Evaluation Parameters for Different Machine Learning Models

Algorithm	Evaluation Parameters					
	AUC	***CA***	***F1***	***Prec.***	***Recall***	***MCC***
KNN	54	33	31	31	33	4
Random Forest	56	37	34	33	37	7
Naïve Bayes	60	36	36	36	36	12
Ada Boost	55	35	32	31	35	5
SVM	57	35	33	33	35	7
Gradient Boosting	58	38	34	34	38	10
XGBooost	58	38	35	40	35	10
Neural Network	55	34	33	32	34	7
Decision Tree	52	32	32	32	32	6
Logistic Regression	59	40	35	35	40	11

The Naïve Base algorithm has slightly better evaluation parameters out of the nine classification algorithms.

Table 5: Confusion Matrix for Naive Base Algorithm

	Predicted					
Actual	**High Impact**	**Less Impact**	**Medium Impact**	**Negative**	**No Impact**	**Total (∑)**
High Impact	57	84	25	24	104	294
Less Impact	69	224	49	48	196	586
Medium Impact	22	46	12	19	35	134
Negative	8	26	14	17	24	89
No Impact	60	108	18	35	257	478
Total (∑)	216	488	118	143	616	1581

As observed from Table 5, there were many underfittings in classifications. The provided confusion matrix evaluates a classification model predicting five impact categories: High Impact, Less Impact, Medium Impact, Negative, and No Impact. The diagonal elements (57, 224, 12, 17, 257) represent correct predictions, while off-diagonal values indicate misclassifications. The model struggles significantly, with an overall accuracy of only 35.9%. High Impact and Negative classes perform poorly, with recalls of just 19.4% and 19.1%, respectively, meaning most true cases are missed.

Medium Impact predictions are particularly weak, with only 12 correct out of 134. Many instances are misclassified as Less Impact or No Impact, suggesting potential class imbalance or model bias toward these categories. This indicates that there is no relationship between the outcome of the last delivery when there is an extra delivery.

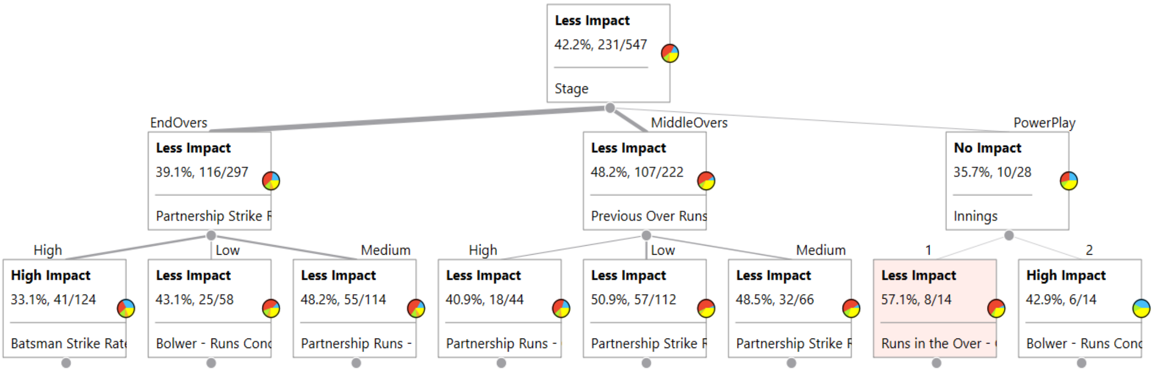

Figure 4: Tree Viewer for Outcome of Delivery

A decision tree viewer was chosen to illustrate the classification results, offering improved readability compared to other algorithms. Figure 4 reveals that while predicting the outcome of an extra delivery remains somewhat uncertain, the most influential parameters are the over number (power play, middle over, or end overs), batsman strike rate, batsman boundaries, and runs scored in the previous over. The figure analyzes the occurrence of "Less Impact" outcomes across different stages, with an overall rate of 42.2% (231/547). Within the End Overs stage, 39.1% (116/297) of instances resulted in "Less Impact," with subcategories like Partnership Strike and Previous Over Runs showing mixed influences (High/Low/Medium impact). Specific scenarios under "Less Impact" varied between 40.9% and 57.1%, though exceptions like "Notional" (35.7% No Impact) and "High Impact" (typo, 33.1–42.9%) suggest overlapping classifications.

This analysis was further refined by selecting specific parameters. For instance, the entire process was repeated exclusively for the power play overs (1-6), but the evaluation metrics remained as poor as those for the entire dataset. This procedure was applied to all variables, yet precision, recall, classification accuracy, and F1 score failed to exceed 50% for any of the classification algorithms.

To identify the most relevant parameters for prediction, a feature selection technique within machine learning was employed using the Azure Machine Learning platform, as shown in Table 6.

Table 6: Evaluation Parameters for Different Machine Learning Models

Feature Selection Technique	**Attribute**	**Contribution**
Pearson Correlation	Ball Number	0.0587
Spearman Correlation	Ball Number	0.0538
Mutual Information	Is Free Hit	0.0638
	Stage	0.0324
	Batsman Strike Rate	0.0159
	Bowler Runs Conceded	0.0143
Chi-Squared	Stage	115.6288
	Batsman Strike Rate	48.7038
	Bowler Runs Conceded	44.5306
	Partnership Stage	44.2802
Kendall Correlation	Ball Number	0.0432
Fisher Score	Ball Number	0.0034

Most feature selection criteria revealed minimal contributions, but the Chi-squared feature selection indicated stronger associations. Specifically, the Chi-squared technique highlighted the match stage, batsman's strike rate, and the number of runs conceded by the bowler at the time of the extra delivery as major contributing factors to the outcome of the extra delivery.

As shown in the Table 7, evaluation parameters were improved once the feature selection was performed using Chi-Squred technique. After performing feature selection, the models show a slight improvement in performance across several evaluation metrics. The Random Forest classifier maintains its lead with an AUC of 60, Classification Accuracy (CA) of 63, and F1-score of 60, while also achieving the highest Matthews Correlation Coefficient (MCC = 5), indicating better-balanced predictions. XGBoost and Naïve Bayes also perform well, with CA and Recall values around 63-64, suggesting better true positive detection. However, some models like AdaBoost and Decision Tree still struggle, with negative MCC values, indicating poor class separation. Logistic Regression has the highest CA (69) but a low Precision (47), meaning it predicts many false positives.

Table 7: Evaluation Parameters for Different Machine Learning Models After Feature Selection

Algorithm	**Evaluation Parameters**					
	AUC	***CA***	***F1***	***Prec.***	***Recall***	***MCC***
KNN	55	62	60	58	62	3
Random Forest	60	63	60	60	63	5
Naïve Bayes	57	63	60	58	63	3
Ada Boost	57	60	56	54	60	-5.6
SVM	53	66	58	58	66	1
Gradient Boosting	58	63	59	58	63	1
XGBooost	57	64	60	60	63	1
Neural Network	59	64	58	57	64	-1
Decision Tree	58	63	57	55	64	-4
Logistic Regression	57	69	56	47	69	-3

6. Conclusion and Future Works

This research corroborates the findings of previous descriptive analytics, demonstrating no correlation between the selected parameters and the outcome of the last delivery when an extra delivery occurs during an over. Both classification and feature selection techniques support this conclusion. Decision tree analysis identified the match stage, batsman strike rate, and runs conceded by the bowler at the time of the extra delivery as significant factors influencing the outcome of the last delivery. This finding was further confirmed by the Chi-Squared feature selection technique.

Future research will utilize the formative measurement model of SmartPLS 4 to investigate and potentially challenge the prevailing expert opinion. Further, CNN models can be utilized after more data collection is performed. As cricket data is continuously generated, subsequent research will benefit from an even larger dataset.

7. References

Ajibade, S.-S. M., Ahmad, N. B. B., & Zainal, A. (2021). Analysis of Metaheuristics Feature Selection Algorithm for Classification. In A. Abraham, T. Hanne, O. Castillo, N. Gandhi, T. Nogueira Rios, & T.-P. Hong (Eds.), Hybrid Intelligent Systems (pp. 214–222). Springer International Publishing.

Anuraj, A., Boparai, G. S., Leung, C. K., Madill, E. W. R., Pandhi, D. A., Patel, A. D., & Vyas, R. K. (2023). Sports Data Mining for Cricket Match Prediction. International Research Journal of Modernization in Engineering Technology and Science, 668–680.

Asanka, P. D. (2025.). Github-Dinesh Asanka. https://github.com/dineshasanka/Cricket-T20-ExtraDelivery/

Asanka, P. D. (2014). Outcome of the Extra Delivery in Cricket—Data warehousing and Data Mining Approach. International Journal of Engineering Research and Technology (IJERT), 03(10).

Asanka, P. D. (2023). Is the Extra Delivery in Cricket as Costly as Experts Think: Descriptive and Diagnostic Analysis on T20 Cricket. International Conference on Applied Sports.

Bhattacharjee, D., & Saikia, H. (2023). Visualizing bowling performance in cricket using contour plot. MOJ Sports Medicine, 6(3).

Callaghan, S. J. (2018). The mechanics of front leg loading during cricket fast bowling: Delivery variations, spell demands, and the effects of strength training [PhD Thesis, Edith Cowan University]. https://ro.ecu.edu.au/theses/2062/.

Dasgupta, I. (2016). Predicting outcome of matches using pressure index: Evidence from Twenty20 cricket. Journal of Sports Science & Medicine, 15(2), 361.

ESPNCricinfo. www.espncricinfo.com/(2025). Duckworth-Lewis Method - Frequently Asked Questions,. http://static.espncricinfo.com/db/ABOUT_CRICKET/RAIN_RULES/DL_FAQ.html .

Gautam, A. K., & Bansal, A. (2022). Impact Analysis of Feature Selection Techniques on Cyberstalking Detection. I-Manager's Journal on Image Processing, 9(4), 21–34. https://doi.org/10.26634/jip.9.4.19138.

International Cricket Council. (n.d.). The Three Formats of Cricket.

Kelly, A. L., Brown, T., Reed, R., Côté, J., & Turnnidge, J. (2022). Relative Age Effects in Male Cricket: A Personal Assets Approach to Explain Immediate, Short-Term, and Long-Term Developmental Outcomes. Sports, 10(3), 39. https://doi.org/10.3390/sports10030039.

Kottearachchi, S. S. S., Jayasinghe, C. L., & Silva, R. M. (2022). An investigation of survival abilities of opening batsmen in One-Day international Cricket. Studies of Applied Economics, 40(1). https://doi.org/10.25115/eea.v40i1.7045.

Kumar, S., Prithvi, H., & Nadini, C. (2022, September). Data Science Approach to Predict the Winning Fantasy Cricket Team—Dream 11 Fantasy Sports.

Mathew, S. M., Mishra, A., & Kumar, M. (2023). Interdisciplinary Application of Machine Learning, Data Science, and Python for Cricket Analytics. In Advanced Interdisciplinary Applications of Machine Learning Python Libraries for Data Science. IGI Global.

Microsoft. (n.d.). Microsoft Machine Learning Studio (classic). Microsoft Power BI (June 2024). (2024). [Computer software]. https://powerbi.microsoft.com/

Orange Data Mining. (n.d.). Data Mining Fruitful and Fun: Open source machine learning and data visualization.

Pramanik, Md. A., Hasan Suzan, Md. M., Biswas, A. A., Rahman, M. Z., & Kalaiarasi, A. (2022). Performance Analysis of Classification Algorithms for Outcome Prediction of T20 Cricket Tournament Matches. 2022 International Conference on Computer Communication and Informatics (ICCCI), 01–07. https://doi.org/10.1109/ICCCI54379.2022.9740867.

Pussella, P., Silva, R. M., & Egodawatta, C. (2023). In-game winner prediction and winning strategy generation in cricket: A machine learning approach. International Journal of Sports Science & Coaching, 18(6), 2216–2229. https://doi.org/10.1177/17479541221119738.

Rahman, Md. M., Shamim, Md. F., & Ismail, Sk. Md. (2018). An analysis of Bangladesh one day International Cricket Data: A machine learning approach. 2018 International Conference on Innovations in Science, Engineering and Technology (ICISET).

Saikia, H., & Bhattacharjee, D. (2010). On the Effect of Home Team Advantage and Winning the Toss in the Outcome of T20 International Cricket Matches. Assam University Journal of Science and Technology, 6(2), 88–93.

Saraswat, D., Dev, V., & Singh, P. (2018). Analyzing the performance of the Indian cricket team using Weighted Association Rule Mining. 2018 International Conference on Computing, Power and Communication Technologies (GUCON).

Shakil, F. A., Abdullah, A. H., Momen, S., & Mohammed, N. (2020). Predicting the Result of a Cricket Match by Applying Data Mining Techniques. In R. Silhavy, P. Silhavy, & Z. Prokopova (Eds.), Software Engineering Perspectives in Intelligent Systems (pp. 758–770). Springer International Publishing.

Silva, D. D., Silva, R., & Jayasinghe, C. (2022). A study of batting out-of-form in One-Day international Cricket. Studies of Applied Economics.

Silva, R. (n.d.). A study of the powerplay in one-day Cricket—Simon Fraser University. A Study of the Powerplay in One-Day Cricket.

Top-10 Most Popular Sports in The World 2023. (2023, July 4). https://sportytell.com/sports/most-popular-sports-world/

Usman-Hamza, F. E., Atte, A. F., Balogun, A. O., Mojeed, H. A., Bajeh, A. O., & Adeyemo, V. E. (2019). Impact of feature selection on classification via clustering techniques in software defect prediction. Journal of Computer Science and Its Application, 26(1). https://doi.org/10.4314/jcsia.v26i1.8

Wickramasinghe, I. (2022). Applications of Machine Learning in cricket: A systematic review. Machine Learning with Applications, 10, 100435. https://doi.org/10.1016/j.mlwa.2022.100435