

Article

RunsGuard Framework: Context Aware Cricket Game Strategy for Field Placement and Score Containment

Aatif Hussain *, Shazia Arshad and Awais Hassan

Department of Computer Science, University of Engineering and Technology, Lahore 54000, Pakistan; shazia.shoaib@uet.edu.pk (S.A.); awais.hassan@uet.edu.pk (A.H.)

* Correspondence: aatif@uet.edu.pk

Abstract: Sports analytics utilizes data analysis techniques and computational methods to gain insights, make informed decisions, and facilitate improvements in the performance of individuals and teams. Cricket is one of the most popular games and continues to evolve worldwide. The availability of ball-by-ball data demands in-depth investigation of player strategies, team dynamics, and the impact of contextual variables. Existing studies explored various aspects of cricket analytics, including detecting key events, predicting outcomes, and ranking teams. However, the literature lacks a comprehensive integrated framework that processes unstructured sports commentary, extracts actionable insights, conducts a thorough player analysis, and develops strategic plans while considering contextual factors. This work aims to propose a bowling and fielding strategy to contain a batsman. For this purpose, we developed a comprehensive context-aware framework that collects data, extracts insights from commentary, identifies player strengths and weaknesses, and proposes cricket bowling and fielding strategies according to the given context. To evaluate this work, we implemented a case study that simulated different scenarios, and our framework suggested bowling and fielding strategies. In these simulations, the proposed strategies consistently demonstrated a substantial reduction in the number of runs that were scored. On average, the framework reduces the batsman's score rate by 33%. These findings underscore the practical effectiveness of research in optimizing field placement and effectively reducing scoring opportunities. Finally, by bridging the gap between data analytics and cricket game strategy, this methodology provides a competitive advantage to coaches, captains, and players. In the future, we aim to involve temporal patterns to understand the evolving behavior of players.



Citation: Hussain, A.; Arshad, S.; Hassan, A. RunsGuard Framework: Context Aware Cricket Game Strategy for Field Placement and Score Containment. *Appl. Sci.* **2024**, *14*, 2500. <https://doi.org/10.3390/app14062500>

Academic Editors: Łukasz Rydzik, Tomasz Palka and Tadeusz Ambrozy

Received: 31 January 2024

Revised: 23 February 2024

Accepted: 1 March 2024

Published: 15 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Advancements in technology and data analytics significant by changes the way that sports are being viewed and engaged [1]. Sports technologies are expected to reach 41.8 billion dollars by 2027 [2], and sports analytics is estimated to be a \$8.4 billion industry by 2026 [3]. It plays a crucial role in sports, embracing performance analysis, real-time tracking, advanced analytics, precise time measurement, data collection through wearables, result-driven communication, instant replay enhancements, biometric data monitoring, virtual training, and AI integration for predictive analysis and facilitates decision-making [4,5].

Cricket is the second-most-followed sport worldwide, with an extensive fan base exceeding 2.5 billion. The International Cricket Council boasts membership in 108 countries comprising 12 full members and 96 associate members. The ICC Men's T20 World Cup in 2022 recorded remarkable engagement with 6.58 billion video views, attracted 78.4 million web and app visitors, and accumulated an impressive 365 million digital streaming viewing hours [6]. Viacom18 secured domestic digital rights for the Indian Premier League's (IPL)

cricket matches through a five-year agreement valued at US \$3 billion, underscoring the substantial economic influence of this sport [7].

Over the past 140 years, cricket has embraced innovation, including technology integration, which benefits coaches and players in game preparation [8]. The increasing popularity of cricket and growing availability of ball-by-ball big data demand more powerful tools and research in cricket analytics [9]. Effective game analysis and planning, including well-planned fielding and bowling strategies, can significantly impact game outcomes by increasing wicket-taking opportunities, restricting runs, and exploiting player dynamics, strengths, and weaknesses, as well as pitch conditions and match contexts [10]. Algorithms for an effective plan require the consideration of pre-game knowledge, real-time knowledge, outlier sensitivity, and reliability estimation in context-aware environments. Furthermore, it is difficult to transform unstructured data into organized information at the activity level of sports. To extract activity-level information, two data streams are typically utilized: video data and text commentary. Extracting activity-level information from video content requires a high computational power [11]. However, text commentary is computationally more efficient, credible, easily available, provides minute delivery specifics, and is storage-friendly, thereby enhancing the depth and accuracy of analysis [11].

Studies [12–16] focused on extracting information from commentary data for multiple purposes, including detecting key events [12], evaluating player performance [13,14], exploring cricket analytics [15], and extracting strength and weakness rules [11,16].

In addition, previous studies [5,9,17–22] collectively addressed various aspects of cricket analytics, encompassing the learning strength and weakness rules of players [17], extracting regions of batsmen through text-commentary analysis [18], machine learning-based team selection [19], and understanding the impact of contextual factors on team performance in T20 cricket through an interpretable machine learning approach [20]. Furthermore, studies have explored the utilization of big data for cricket match outcome prediction [9], sports analytics for cricket games [21], offering a unique approach to ranking cricket teams [23], and conducting a comparative analysis of pitch ratings for all formats of cricket [22]. Finally, studies [24–27], collectively explored strategic decision-making and performance evaluation in cricket, covering optimal playing strategies against specific bowling types using game theory [24], player-aware resource compensation in interrupted matches [25], investigating the influence of the T20 cricket on Test cricket performance and team quality [26], and estimating a batsman's shot selection in T20 cricket to guide strategic decisions related to fielder placement [27].

With reference to cricket analytics, the literature lacks the following:

1. In-Depth Sport-Specific Attributes: Many existing studies depend on aggregate data [12,13,17,28], yet a strategic plan demands more nuanced and in-depth information. This plan requires the incorporation of the actual execution of sports activities, encompassing the essential sport-specific attributes inherent in these activities.
2. Limited Comprehensive Frameworks: Existing research lacks a comprehensive integrated framework for extracting actionable insights from unstructured ball-by-ball commentary, conducting thorough player analysis, and developing strategic plans that consider contextual factors.
3. Context-Aware Strategies: A gap exists in the understanding and implementation of context-aware strategies in cricket analytics. Current approaches [24–27] do not adequately consider contextual factors such as powerplays, the number of fielders in the circle and outer circle, and specific line and length variations.

Problem Statement: Building upon the preceding discussion, there is a need to propose an integrated framework to extract actionable insights from unstructured data (ball-by-ball commentary), identify the strengths and weaknesses of a player, and develop fielding and bowling strategies while considering contextual factors.

Specifically, this study seeks to address three questions.

- Q1: How can unstructured commentary text data be processed and represented as actionable knowledge?
- Q2: How can an in-depth player analysis be conducted to identify strengths and weaknesses in the context of cricket?
- Q3: How can context-aware fielding and bowling strategies be tailored while considering individual player attributes?

To answer these questions, this study introduces a comprehensive framework (RunsGuard) to develop a cricket game strategy in the context of contemporary competitive dynamics and individual player idiosyncrasies. To effectively restrict scoring capabilities, the solution offers context-aware measures that include bowling and fielding placement tactics designed for specific power-play situations, variations in the number of fielders in the inner and outer circles, and optimization of the delivery line and length. The proposed approach provides field placement and bowling strategies by learning a player's historical approaches and frequent moves against a specific batsman, considering match-specific contexts such as powerplays and fielding constraints. It focuses on factors such as field positions, shot positions, delivery specifics, and bowling tactics to strategically place fielders based on a batsman's strengths and weaknesses in scoring zones, thereby minimizing scoring opportunities and enhancing team performance.

In Section 2, we explore the existing literature review on cricket player analysis, discuss relevant methodologies and techniques used in similar studies, and outline our proposed methodology. Section 3 delves into the chosen methodology and Section 4 presents the results and discussion. Section 5 presents limitations and future directions, and finally, Section 6 presents conclusions.

2. Literature Review

In the pursuit of crafting a robust and precise framework for optimizing performance in cricket game strategy, this section summarizes 32 relevant research works.

2.1. Structuring Unstructured Text Commentary

Miraoui et al. [12] employ natural language processing to automate the identification of crucial actions in sports events, categorizing them through systematic analysis of live commentaries from diverse sources. To improve the identification of significant events and offer more complex insights into athletic narratives, this study investigated sentiment analysis. Association rule mining was used by Arif et al. [13] to extract features, such as bowling position, location, and wickets per inning. Using statistical techniques to examine post-text extraction, this study seeks to provide meaningful data. The predicted results have the potential to improve the cricket team selection strategy and assist squad selectors in selecting the best squad combinations for given situations. An in-depth analysis of cricket player performance was provided by Goel et al. [15], utilizing NLP natural language processing, sentiment analysis, topic modeling, and NER on commentary material. This thorough examination, which results in a player impact model, reveals the hidden characteristics of the performance. Sports analysts, coaches, players, and spectators can benefit from this study, which deepens our knowledge of sports.

A real-time approach for analyzing cricket comments is presented by Roy et al. [14], extracted performance indicators, and suggested the best players for forthcoming matches. A comprehensive evaluation confirms the effectiveness of the structure, propelling cricket analytics forward and developing a dependable method for commentary-based real-time player assessment. Behera et al. [11] neglected the short text commentary dataset, which is renowned for being concise and sometimes disregarded by the machine learning community, to extract strength and weakness rules. The research gathered comprehensive player data illustrated by domain-specific characteristics. The researchers created a semantic relationship between bowlers and batters by using correspondence analysis for discrete random variables. This relationship is displayed through biplots, which allow the derivation of human-readable strength and weakness guidelines.

2.2. Cricket Player Analysis: Strengths and Weaknesses

Behera et al. [17] and Khan et al. [18] focused on evaluating the strengths and weaknesses of cricket athletes. Behera et al. employed association rule mining on player statistics, such as runs, boundaries, and innings, to provide insights for scouts and coaches in team decision making. Khan et al. presented a text-mining method to extract strong scoring shot points and identify challenging shot regions for batsmen. They proposed a mechanism to calculate region-wise strike rates using the T20 cricket text commentary from espncricinfo. This assists coaches, players, and opposing bowlers in making informed decisions based on batsmen's strengths and weaknesses.

Shetty et al. [19], Biswas et al. [5], and Mody et al. [28] focus on player performance. Shetty et al. addressed the problem of selecting the best playing 11 among Indian cricket players, considering variables such as surface type, opponent, and ground. Batsmen, bowlers, and all-rounders are predicted with data of one-day internationals and have been predicted with good accuracy through the application of the Random Forest algorithm. Biswas et al. have applied the linear regression, support vector machine, random forest, naive Bayes for the prediction of runs by batsmen and performance of the bowlers in cricket sports analytics. In this regard, their study indicates that the machine-learning potential of player selection and strategic planning before future contests. Mody et al. presented an artificial neural network (ANN) based on a prediction model for the performance of cricket players with regards to bowling and batting in Twenty20 matches. The study had concluded that an Artificial Neural Network (ANN) offers higher accuracy as compared with other classifiers, such as Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM).

Puram et al. [20] used machine learning methods, with analysis provided in the forefront of Bayesian additive regression tree (BART), to study factors acting upon T20 cricket team scores and contextual variables. This research has identified key performance-influencing factors, which will help T20 teams identify the most effective pre-game and pre-season strategies. The authors also state that the study is not without its own limitations, which involve possible data limitations and the need for additional confirmation in different cricketing contexts. Tirtho et al. [29] synthesized multi-objective optimization and machine learning (ML) for predicting players in upcoming tournaments of cricket. This paper provides a technique of picking up the cricketers using the multi-objective optimization method and applying machine learning techniques. It helps one to spot possible weaknesses, gather information, and validate in real-world cricket situations.

Mahbub et al. [30] and Pussella et al. [31] focused on cricket match forecasting. The most used techniques for this study were the Support Vector Machine (SVM), Naive Bayes, and Random Forest (RF) in machine learning to predict the top 11 cricket players for Bangladesh in One-Day International (ODI) matches. For batsmen, they had derived 94% precision, and for bowlers, 93% of the time. The study by Pussella et al. utilized a dataset with information drawn from 812 Indian Premier League games to predict winners of T20 cricket matches through LASSO algorithms. The accuracy obtained was in the range 53.08% to 97.65% with the models Naive Bayes (NB), Logistic Regression (LR), and Support Vector Machines (SVM). In other works, they also introduced a winning strategy generator and decision making through web-based dynamic programs. Awan et al. [32] report the case of integration of machine learning (ML) into big data analytics in the game of cricket for the prediction of players and the match outcome. The team scores were predicted with linear regression models: with and without big data processing with Spark ML.

The results indicate promising accuracy and suggest potential applicability in other sports. However, limitations may include data volume requirements and generalizability of the models to different cricket formats. Kapidia et al. [21] employed machine learning (ML) to predict cricket-match outcomes in the Indian Premier League (IPL) based on historical data. The study identifies Random Forest as the top-performing algorithm for the home team advantage feature set. However, no algorithm demonstrates a strong performance for the toss-featured subset. This study offers valuable insights into IPL match outcome

prediction using ML techniques; however, potential limitations may arise in accurately predicting toss-related outcomes.

Das et al. [23] offer a unique approach to ranking cricket teams that takes into account both non-cricketing and a variety of performance criteria. It surpasses the current methodology and ICC rankings in the Test, One Day International (ODI), and Twenty20 (T20) formats and generates the Commitment Index Rank (CIR) for teams. Kumar et al. [22] proposed a special technique for evaluating the pitch consistency in cricket matches played in international tests. It includes games played between March 2017 and March 2019, omitting the seasons impacted by COVID-19. For goal pitch evaluation, the PROMETHEE II method was applied with measurable parameters such as runs/day, wickets/day, runs/over, runs/wicket, and overs/wicket. Pitch rating consistency and two unique rating clusters are shown in the results, which bode well for further research and creative statistical methods for sports pitch analysis.

Sarangi et al. [33] offered an approach that examines predictors of success in one-day international cricket matches for four international teams. Binary logistic regression analysis was used in this study. The analysis shows that fielding dismissals and the bowler economy have a considerable impact on the chances of winning for each of the four teams. By combining real-time statistics with historical trend data, the proposed models improve victory rates and provide insightful information for coaches and team managers.

Sajib et al. [34], Pramanik et al. [35] and Krishnan et al. [36] focus on cricket match performance. Sajib et al. used ESPNcricinfo data to analyze the key variables influencing the performance of a Bangladesh men's cricket team. Using binary logistic regression, they identified location, most wickets, and one-digit scores as significant factors. In their assessment of non-ensemble and ensemble classifiers, Pramanik et al. forecast the results of the Bangladesh Premier League (BPL) Twenty20 matches. The dataset includes pre- and post-match characteristics for the 2011–2012 through 2019–2020 seasons. According to their results, the Gradient Boosting technique achieves 93.39% accuracy in comprehensive forecasts, whereas the (KNN) obtains 64.58% accuracy in pre-match forecasts. Krishnan et al. used machine learning to provide information about a team's performance for the benefit of franchisees, investors, and spectators. Logistic Regression, Decision Tree, and Random Forest algorithms were used to analyze IPL matches from 2008 to 2022 and forecast the team winning probability.

2.3. Context-Aware Strategies

Game theory is used by Das et al. [24] to investigate the best batting tactics in cricket against various bowling styles. In light of the batsman's critical role in shaping the posture of the field team, this study attempts to reveal practical batting strategies for different types of bowlers through real-time data analysis. However, participant variability and complexity of practical gaming environments may have major drawbacks. Zia et al. [25] describe a dynamic player-aware resource compensation model (PRCICM) for handling interrupted cricket matches. Through resource allocation based on the historical usage of squads, athlete engagement, and match location, this approach adapts to different squad compositions and playing tactics.

Nikholls et al. [26] note an increase in sixes, a decline in Test draws, and an increase or decrease in run rates during their examination of the influence of T20 cricket on Test cricket metrics in the leading ICC Test teams during 2000 and 2020. This research emphasizes the importance of six-hitting percentages for players to raise their scoring capabilities in T20. Weaknesses include the emphasis on Test cricket, the recommendation for more study on the impact of T20 matches in one-day internationals, and the strategic consideration of variables such as home-field advantage and team caliber. In six different cricket field settings, Das et al. [27] evaluated a batsman's stroke selection against a variety of bowling styles using a probabilistic model. For shot probability, the research used a multinomial distribution; for transition probabilities and mean recurrence time, it used Markov chain analysis. To

help fielding team captains make tactical judgments about fielder placement and bowler selection against particular batters from rival teams, this study offers insightful information.

Studies [37,38] have presented innovative approaches in their respective fields. Y. Yang et al. [38] optimizes data freshness and privacy in mobile crowdsensing using mean field game theory and encryption techniques. Yang et al. [37] used stochastic geometry to assess semantic keyword coverage in text, offering a new perspective on the semantic sensing performance. Both studies contribute valuable theoretical frameworks and practical insights into optimizing the network performance and data analysis. Rahimian et al. [39] apply deep reinforcement learning to study soccer match data. They provided strategies for optimal in-game decisions. Attackers recommend long-distance shots. For defenders, they suggest tailored actions based on pitch location. The study [40] introduced a unique method for choosing fantasy cricket teams. It uses data analysis and a genetic algorithm to select better teams than the traditional methods. This approach was tested using the Dream11 app, which showed better results. It fills a gap in the research on fantasy cricket team selection.

Anuraj et al. [41] and Subramanian et al. [42] focus on cricket match prediction. Anuraj et al. used sports data mining to forecast cricket matches, particularly the T20 International World Cup matches. Real-world information, such as location, toss winner, team standings, and previous wins against particular opponents, shows how successful their forecasting methods are. Subramanian et al. presented a prediction system for ongoing cricket matches, considering player quality, team form, toss, home benefit, and match time. A total of 187 ODI matches between 2015 and 2017 were evaluated, and the dynamic logistic regression model demonstrated strong forecasting accuracy after adjusting for various factors.

In conclusion, there are notable gaps in the existing literature. First, there is a need for a more profound exploration of sport-specific attributes (ball-by-ball attributes, such as line, length, shot type, position, and ball no. etc.), moving beyond reliance on aggregate data. Second, the absence of a comprehensive integrated framework hinders the extraction of actionable insights from unstructured ball-by-ball data such as text commentary, comprehensive player analysis, and strategic planning. Finally, there is a discernible gap in understanding and implementing context-aware strategies in cricket analytics.

3. Proposed Framework for Cricket Game Strategy and Field Placement Optimization

This section provides a detailed overview of the RunsGuard framework proposed for this research, delineating its utilization from data acquisition through strategic planning to enhance performance in cricket matches. The proposed framework depicted in Figure 1 comprises of four major steps that illustrate the overall architecture. The first step (Section 3.1) is to obtain ball-by-ball commentary data from various sources and compile them through cleaning, normalization, and segmentation. For our study, text commentary data were carefully selected from two major sources, ESPNcricinfo [43] and Cricbuzz [44] based on a set of well-defined criteria to ensure the robustness and relevance of our analysis. The criteria include:

- Relevancy: Ensuring that commentary data closely reflects match events and player actions.
- Comprehensiveness: Ensuring coverage of every ball across a wide range of matches.
- Accuracy: Relying on known sources for accurate reporting.
- Availability: Prioritizing data sources that offer extensive historical archives.
- Timeliness: Including recent and historical matches.
- Consistency: Maintaining a consistent format and level of detail.
- Diversity: Covering different leagues, tournaments, and match situations.

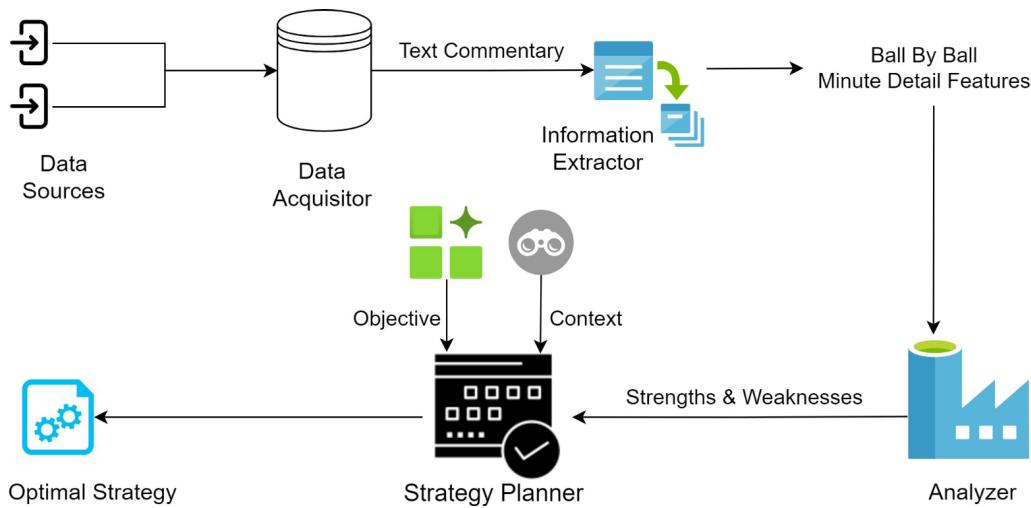


Figure 1. Overall Architecture of the Framework.

By adhering to these criteria, our selection of text commentary data aims to provide a foundation for a detailed analysis, contributing valuable insights to the field of sports analytics.

The second step (Section 3.2) involves the extraction of information using natural language processing (NLP) techniques with conditional random fields (CRF) on the compiled commentary data. The third step (Section 3.3) is a pattern and trend inspection stage, which is carried out by the analyzer to inspect the features and translate them in terms of the strengths and weaknesses of a player. The performance index (PI) in Equation (4) given here is computed in this step based on the total runs scored with specific identified patterns and the frequency ratio (FR) stated by Equation (3) through the occurrence of these patterns. Finally, Strategy Planner Step (Section 3.4) recommends the placement of fielders for bowling strategies based on PI, FR, constraints, and contextual factors using a mathematical model. This model aims to minimize PI while increasing FR, thereby improving the performance during cricket matches. The operational procedure for each step is as follows:

3.1. Data Acquisitor

The organized collection and compilation of textual commentary data from matches is the main objective of this step. The mechanics of this process are outlined in Algorithm 1.

We used specific tools and technologies to carefully collect and preprocess the collected unstructured commentary dataset, ensuring that it is ready for in-depth analysis. The process begins with the web scraping tool Scrapy, which is designed to navigate and pull information from the HTML source code. This important step pulls out the text commentary from the selected match URLs. This is achieved by focusing on specific HTML tags. These tags were identified by examining the Document Object Model (DOM). After successfully extracting the commentary text, we proceeded to the preprocessing step. In this phase, we relied on Python's handheld tools. Pandas was used for manipulating the data, and a Natural Language Toolkit (NLTK) was utilized for processing the text. These tools assist in deleting unwanted characters, making the data uniform, and segmenting the data into analytical units. Using Python's regular expressions library (re), we fixed special characters and formatting issues, keeping the commentary clean and consistent. The systematic collection and preparation of text commentary data provides foundations for the next phase, which exposes the useful characteristics of cricket commentary.

Algorithm 1 Input Stream Acquisition

- 1 Input: List of match-urls
 - 2 Processing:
 - (a) Web Scraping and Text Commentary Extraction:
 - Retrieve the HTML source code for the given match URLs.
 - Analyze the Document Object Model (DOM) structure of the source code.
 - Parse the HTML source code to get desired HTML tags.
 - Extract the text commentary from DOM.
 - (b) Preprocessing:
 - Store the raw text commentary.
 - Clean the data (unnecessary characters and formatting).
 - Handle special characters.
 - Normalize the commentary data.
 - Segment the commentary into relevant sections for further processing.
 - (c) Validation:
 - Validate accuracy by cross-referencing with other sources.
 - Fix the issues.
 - 3 Output: Preprocessed commentary data
-

3.2. Information Extractor

In this step, valuable features are extracted from the preprocessed text commentary. Figure 2 illustrates the procedure. The commentary data we utilized is a detailed textual representation of the live ball-by-ball activity during a cricket match. These data do not reflect the commentator's personal opinions but rather provide a factual account of events as they unfold in the field. Our data extraction process is designed to parse this textual information to retrieve only objective data points that are crucial for our analysis, such as the bowler's name, batsman, outcome of the delivery, delivery length, delivery line, shot placement, and shot type.

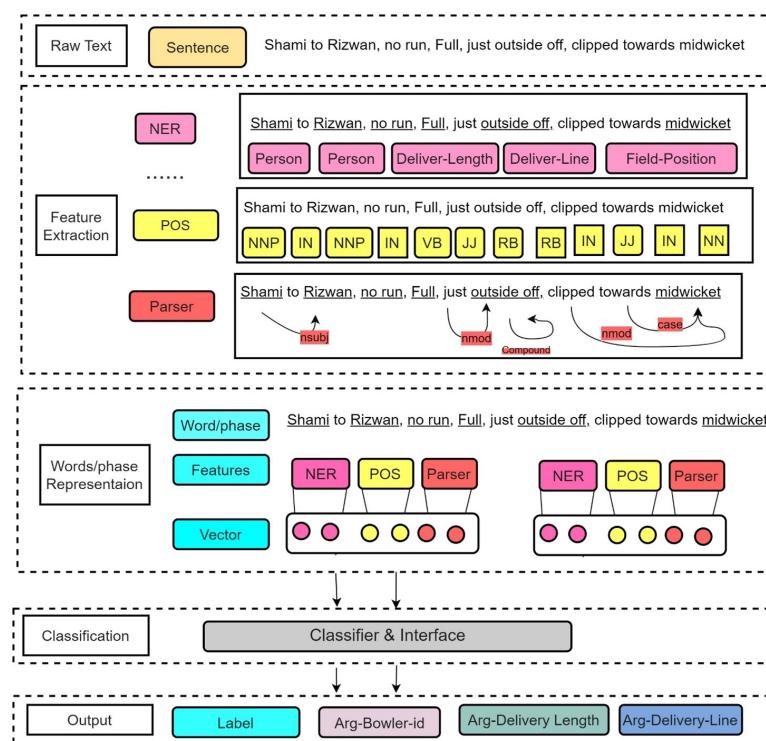


Figure 2. Demonstration of Feature Extraction from Text Commentary.

We used the NLP spaCy library for its powerful NLP capabilities to tokenize textual expressions. This library enabled us to identify key features of speech, such as part-of-speech tags and word embeddings, which are important for our in-depth analysis. This enables us to systematically classify text commentary into predefined features, such as Delivery Line, Delivery Length, Bowler, Batsman, outcome shot type, and others, to understand the game's dynamics. Owing to its concise and somewhat structured format, a cricket commentary is easier to analyze than unstructured text. The conditional Random Field (CRF) model was used to predict game-specific labels using the sklearn-crfsuite. We opted to use CRF simply because it is good in understanding word sequences and context, which form the biggest chunk in reading cricket commentary. The CRF model is trained with expert-checked commentary and is made sure that the model learns and can correctly identify the key elements of the game, the actions players take, and the results of these actions. We fine-tuned the CRF model settings after training to obtain the most accurate results. The extracted attributes serve as the foundation for initiating an in-depth analysis of player performance in the next step. A detailed breakdown of this process is presented in Algorithm 2.

Algorithm 2 Information Extraction

- 1 Input: Text commentary dataset
 $X = \{x_1, x_2, x_3, \dots, x_n\}$ where x_i represents the textual commentary instances.
- 2 Output: Sequences of features of each commentary
 $Y = \{y_1, y_2, y_3, \dots, y_n\}$ where y_i is a sequence of features corresponding to x_i .
- 3 Processing:
 - (a) Tokenize the commentary.
 - (b) Extract linguistic features (POS tags, word embeddings).
 - (c) Define feature categories (Delivery Line, Length, Speed, Field Position, Shot Type, etc.).
 - (d) Generate labeled training data with token-level labels.
 - (e) Train a Conditional Random Field (CRF) model.

$$P(Y|X) = \frac{1}{Z(X)} \exp \left(\sum_{i=1}^N \sum_{j=1}^L (\lambda_j f_j(y_i, y_{i-1}, X, i)) \right) \quad (1)$$

where:

- $P(Y|X)$ is conditional probability.
- $Z(X)$ is a normalization factor.
- λ_j model parameters that assign weights to the feature functions.
- $f_j(y_i, y_{i-1}, X, i)$ refers to characteristic functions evaluating label alignment with input order.
- The trained CRF model extracts features and tags sequences in new commentary.

- (f) Evaluate the model using validation data.

- 4 Inference Phase:

- (a) For all new commentary data:
 - (b) Tokenize the commentary text.
 - (c) Extract linguistic features.
 - (d) Predict the labels for each token using the trained CRF model, as shown in Equation (1).
 - (e) Map labels to corresponding feature categories.
 - (f) Store extracted features for each delivery.
-

3.3. Analyzer Unveiling Player's Strengths and Weaknesses

By analyzing the feature distributions, this step uncovers distinctive trends and patterns that provide insights into a player's strengths and weaknesses across various dimensions of the game. The comprehensive process is provided in Algorithm 3 and illustrated in Figure 3.

Algorithm 3 Analyzer

1 Input: Feature Data (FeatureDataSet)

$X = \{x_1, x_2, x_3, \dots, x_n\}$ where x_i represents the features extracted in the previous step from commentary ball-by-ball data.

2 Processing:

Equation (2) illustrates the sequential steps involved in processing as follows:

$$Y = SW(FCA(FA(P(X)))) \quad (2)$$

where:

- P represents preprocessing.
- FA denotes feature analysis.
- FCA stands for feature combination analysis.
- SW refers to strengths and weaknesses analysis.

(a) Preprocessing

- Handle missing values in the Feature Data (FeatureDataSet).
- Perform data normalization to ensure consistency in Feature Data (FeatureDataSet).
- Apply any necessary data transformations to enhance the dataset (FeatureDataSet).

(b) Feature Analysis

- Perform correlation analysis on the Feature Data to find relationships between features.
- Utilize Principal Component Analysis (PCA) to identify key performance-related features.

(c) Feature Combinations Analysis:

- Calculate the Frequency Ratio for combinations of performance-related features using the formula:

$$FR = \frac{\text{Total occurrences with a specific combination}}{\text{Total occurrences of all combinations}} \quad (3)$$

- Calculate the Performance Index for combinations of performance-related features using the formula:

$$PI = \frac{\text{Total scores with a specific combination}}{\text{Total balls with a specific combination}} \quad (4)$$

(d) Identification of Strengths and Weaknesses Rules (SW):

- Performance index (PI) based categorization
- Frequency ratio (FR) based categorization

3 Output: Identified patterns of strengths and weaknesses (PatternSet), including performance indices and frequency ratios.

$$Y = \{PI, FR, SW_c, y_1, y_2, y_3, \dots, y_n\}$$

where SW_c, y_i is a sequence of key performance-related features.

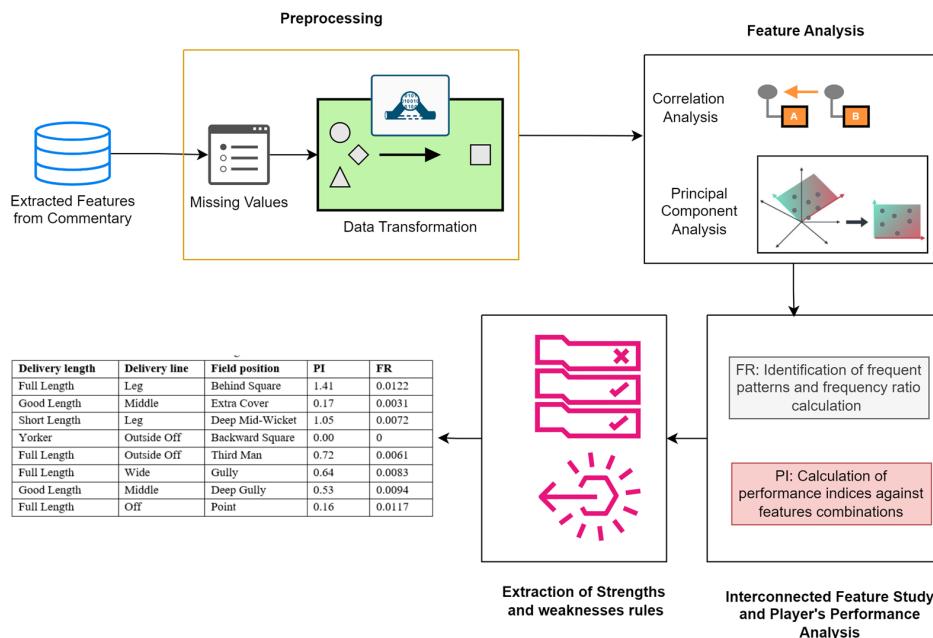


Figure 3. Overall Flow Details of Analyzer Step.

The pivotal objective of this step is to pinpoint the strong and weak zones of the batter to strategically contain runs. To achieve this, we delve into the batsman's performance across various feature combinations, focusing particularly on how frequently a batsman plays shots in specific zones when faced with different lines, lengths, and match contexts (e.g., powerplay). This nuanced analysis is facilitated by calculating two critical metrics: the Performance Index (PI) and the Frequency Ratio (FR), which are derived from detailed ball-by-ball data.

Initially, we prepared the feature data by addressing missing values, normalizing the data to ensure uniformity, and applying the necessary transformations to refine the dataset for analysis. We employed correlation analysis to explore the relationships between different gameplay features, utilizing Python's NumPy library. Principal Component Analysis (PCA), facilitated by the scikit-learn library, was then applied to identify the most significant features that influence a batsman's performance. Feature Combination Analysis involves calculating the PI and FR for specific feature combinations related to the batter. The PI measures the impact of these combinations on scoring, whereas the FR assesses how often the batsman opts for certain shots in given zones. These calculations are crucial for understanding the efficacy of the different batting strategies. Based on the PI and FR metrics, we categorized the identified patterns into the strengths and weaknesses of the batsmen. The calculated metrics, PI and FR, lay the groundwork for optimizing the fielder positions in the subsequent strategy planning step.

3.4. Strategy Planner—Crafting Optimal Strategies

In this phase, by leveraging the insights derived from step 3.3, the process develops a strategy to enhance performance by restricting the capability of the batsman to score runs. The mathematical model below is formulated to optimally allocate fielders in different regions of the ground, considering different factors such as delivery lines, lengths, and contexts. Assignments are indicated by binary variables, and the objective function aims to minimize the weighted aggregate of the performance indices while simultaneously maximizing the frequency ratios. Sets, parameters, and variables are established. Accurate fielders and position allocations are ensured by these constraints. The binary variables indicate assignments, and an optimization algorithm is applied to minimize the objective functions.

In this case, the field placement problem has two objectives.

- Minimize Performance Index (PI): A lower PI suggests that the batter is less efficient at a certain field position. As a result, lowering the PI helps to limit the batsmen's score in that position.
- Maximize Frequency Ratio (FR): A higher FR implies that the batter plays more balls in a specific region. Maximizing FR helps in the deployment of fielders in areas where the batsman often plays shots, thus improving the chances of restricting his/her scoring. Thus, the overall objective function is expressed by Equation (5).

$$\text{Optimization Function} = w_1 \times \text{Minimize PI} - w_2 \times \text{Maximize FR} \quad (5)$$

where w_1 and w_2 are the weights assigned to each term in the objective function. They drive the optimization process and determine the values that result in an optimal solution aligned with the two above-mentioned objectives. The optimization problem expressed by Equation (5) can be mathematically modeled as follows:

3.4.1. Set (S)

F: Set of fielders (excluded bowler and wicketkeeper)
P: Set of field positions
X: All combinations of features, and contextual information
 P_{in} : All inside circle field positions
 P_{out} : All outside circle field positions

3.4.2. Variable (V)

$\text{Allocate}_{fi}(X_k)$: A binary variable that indicates whether fielder f is assigned to position i for combination k .

3.4.3. Constraints (C)

For every combination X_k , a single position is allotted to every fielder.

$$\sum_i \text{Allocate}_{fi}(X_k) = 1 \quad \forall f, \forall k \quad (6)$$

No more than one fielder is assigned to any given position for any combination X_k .

$$\sum_f \text{Allocate}_{fi}(X_k) \leq 1 \quad \forall i, \forall k \quad (7)$$

The fixed number of fielders assigned to both inside and outside the circle.

$$\sum_{i \in P_{in}} \sum_f \text{Allocate}_{fi}(X_k) = N_{in} \quad \forall k \quad (8)$$

$$\sum_{j \in P_{out}} \sum_f \text{Allocate}_{fj}(X_k) = N_{out} \quad \forall k \quad (9)$$

Binary constraints on assignment variables:

$$\text{Allocate}_{fi}(X_k) \in \{0, 1\} \quad \forall f, \forall i, \forall k.$$

3.4.4. Objective Function (Z)

$$\begin{aligned} \text{Min. } Z = & \sum_k \left(\sum_{i \in P_{in}} \text{PI}_{in}(P_i, X_k) + \sum_{j \in P_{out}} \text{PI}_{out}(P_j, X_k) \right) \\ & - \sum_k \left(\sum_{i \in P_{in}} \text{FR}_{in}(P_i, X_k) + \sum_{j \in P_{out}} \text{FR}_{out}(P_j, X_k) \right) \end{aligned} \quad (10)$$

where: $\text{PI}_{in}(P_i, X_k)$: Performance index for inside circle fielders at position P_i for combination X_k .

$PI_{out}(P_j, X_k)$: Performance index for outside circle fielders at position P_j for combination X_k .

$FR_{in}(P_i, X_k)$: Frequency ratio for inside circle fielders at position P_i for combination X_k .

$FR_{out}(P_j, X_k)$: Frequency ratio for outside circle fielders at position P_j for combination X_k .

Algorithm 4 ensures the optimal assignment of fielders inside and outside the circle to minimize the performance index and maximize the frequency ratio across various combinations of delivery lines, lengths, and contextual factors.

Algorithm 4 Planner

1 Input: Identified Feature Combinations from the Analyzer Step

(a) F, X, P, P_{in}, P_{out} as described in Section 3.4.1

2 Processing:

(a) Initialize variables $Assign_{fi}(X_k)$ to zero for all f, i, k .

(b) Define constraints (C) as described in Section 3.4.3 in Equations (6)–(9).

(c) Compute $PI_{in}, PI_{out}, FR_{in}, FR_{out}$.

(d) Use Linear Programming to minimize the defined objective function listed in Equation (10) subject to the given constraints (C).

3 Output:

(a) The outcome yields an optimized fielder assignment strategy that contributes to strategic decision making in cricket games.

4. Results and Discussion

In this study, we employed a case study approach to examine and the results were analyzed. The selection of a representative case study is a fundamental aspect of this research. The following section presents the results derived from this case study, shedding light on the key insights and implications of our research objectives.

4.1. Case Study

We selected player [45] as the focal point for this in-depth case study. The significance of this choice lies in a player's reputation as a formidable and accomplished player. The chosen case study provides a comprehensive foundation for our research, allowing us to delve into the specific details and draw meaningful conclusions.

4.2. Input Data Acquisitor

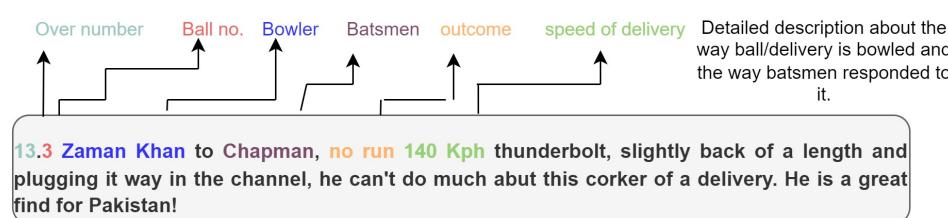
In the case study, match commentary data were sourced from reputable platforms, such as ESPNcricinfo [43] and Cricbuzz [44], serving as the foundational dataset for the analysis. For reference, Table 1 provides a snapshot of the acquired data. The dataset spans all T20I matches from 2012 to 2022, encompassing 27,488 deliveries. Of these deliveries, 2807 correspond to instances in which the prominent player under study was the batsman.

Table 1. Snapshot of Collected Data.

Over No.	Ball No.	Bowler Name	Batsmen Name	Outcome	Total Runs	Comment Cricinfo	Comment Cricbuzz
1	1	Saqib Mahmood	Babar Azam	Four	4	Back of a length outside off, rides the bounce and clubs this through midwicket! Belting start, Curran gives chase but it's agonisingly out of reach, and he can drag it back in with a despairing dive	Babar Azam and Pakistan are underway with a boundary. A harmless short ball marginally outside off, Azam stands tall and pulls it to the right of mid-on. Didn't over-hit and focused on the timing. Curran gives a chase from mid-on and puts in a slide, in vain
1	2	Saqib Mahmood	Babar Azam	Dot	0	Length outside off, driven to cover on the edge of the ring. 86 mph/138 kph	138 kph, fuller than good length, Babar Azam plays a good-looking cover drive, can't find the gap

4.3. Information Extractor

Figure 4 provides an illustrative snapshot of the input commentary. Furthermore, Table 2 presents the extracted features as the outcome of this step, encompassing information on the delivery line, delivery length, and various other elements, which is one of the objectives of this study, as listed in Research Question 1. To ensure the accuracy of our data, we used two reliable sources: ESPNcricinfo [43] and Cricbuzz [44]. We compared the information obtained from these platforms to identify and fix any differences. If needed, we also looked at match videos and commentaries to check our data. This process helped us ensure that our analysis was correct and trustworthy. The presentation of structured data represents a crucial step in transforming unstructured commentary into actionable knowledge.



Text Commentary Example

Figure 4. Sample Commentary.

Table 2. Snapshot of Extracted Features.

Over	Ball No.	Bowler	Batsmen	Outcome	Runs	Length	Line	Shot Type	Position	Hit Type	Variation	Speed	Field Activity
1	1	Saqib Mahmood	Babar Azam	Four	4	Short	Outside Off	Cover Drive	Mid on	Middled	Off Cutter	Slow	Dive
1	2	Saqib Mahmood	Babar Azam	Dot	0	Good Length	Outside Off	Cover Drive	Cover Point	Middled	Normal	138	Normal

4.4. Analyzer—Unveiling Player’s Strengths and Weaknesses

The case study on the selected player’s performance has unveiled intriguing findings through performance index computation for various feature combinations. These findings highlight the strengths and weaknesses of the delivery lines, delivery length, field positions, shot types, and runs. This deep data analysis informs precision cricket strategies, including optimized bowling and fielding tactics aimed at limiting the batsman’s scoring.

Figure 5 shows his shot preferences, with a notable emphasis on the drive and cut shot types; Figure 6 showcases his run distribution, with ‘singles’ being the most frequent outcome; Figure 7 highlights the dominance of deliveries outside the off-line region in his gameplay, and Figure 8 offers a view of his field positioning, revealing how he navigates various field setups.

Player strengths and weaknesses were identified based on their performance against different delivery lines, revealing “outside off” as a strength and a relative weakness against “leg” deliveries Figure 9. The player excelled with “full-length wide” deliveries but showed vulnerabilities to “good-length” deliveries on the off-stump line Figure 10. In addition, strong scoring zones were identified in “Deep Midwicket” and “Long On”, with relative weaknesses in the deep fine leg or deep square leg for specific deliveries Figure 11. The nuanced insights derived from this step are shown in Table 3, which serves as the objective of this study in response to Research Question 2, which seeks to unravel the intricacies of comprehensive player analysis to identify strengths and weaknesses in the realm of cricket. The PI values, calculated based on the delivery length, line, and field position, indicate the player’s scoring ability in specific combinations. Higher PI values indicated strong zones, whereas lower values indicated weak zones. Similarly, a higher FR value indicates a player’s frequent playing zone for a specific combination.

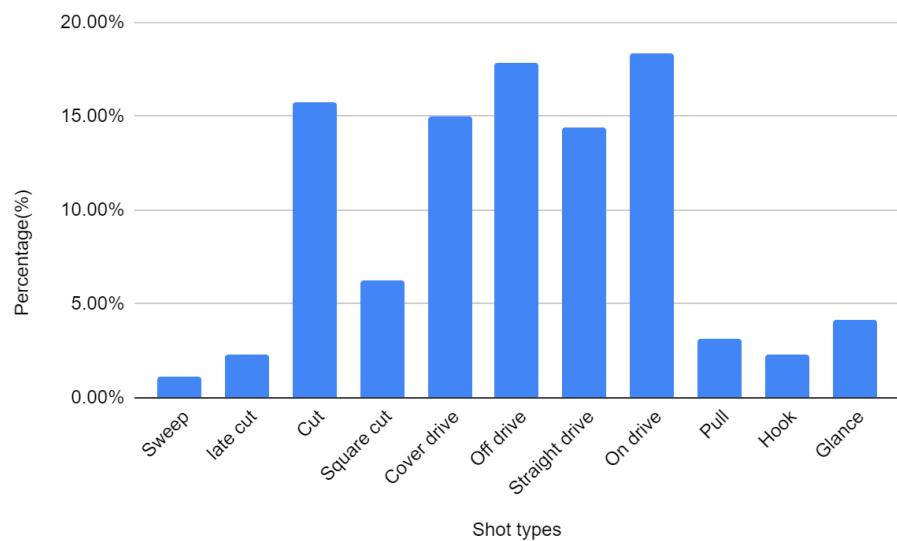


Figure 5. Distribution of Shot Types.

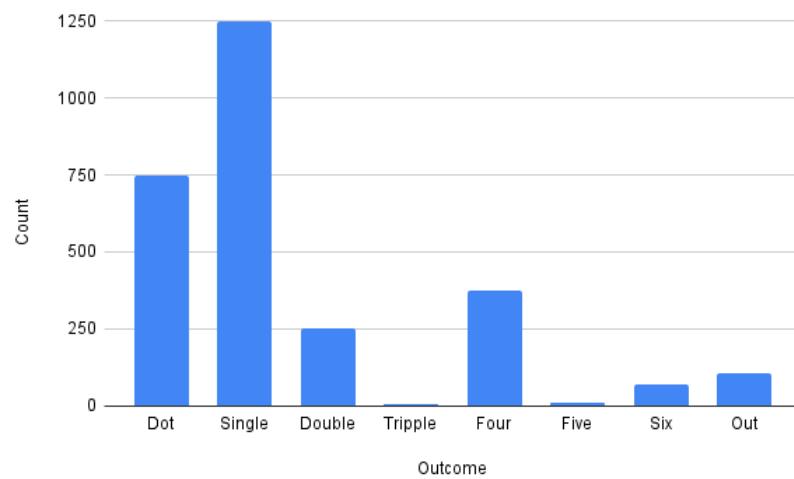


Figure 6. Distribution of Outcomes.

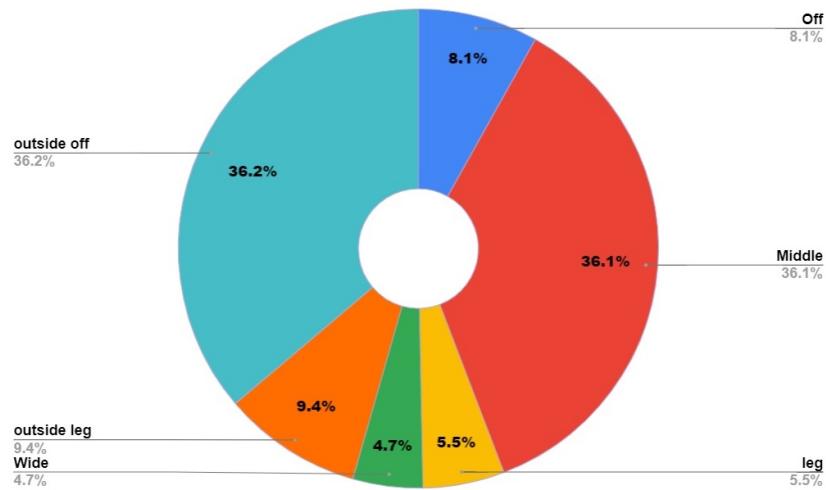


Figure 7. Distribution of line types.

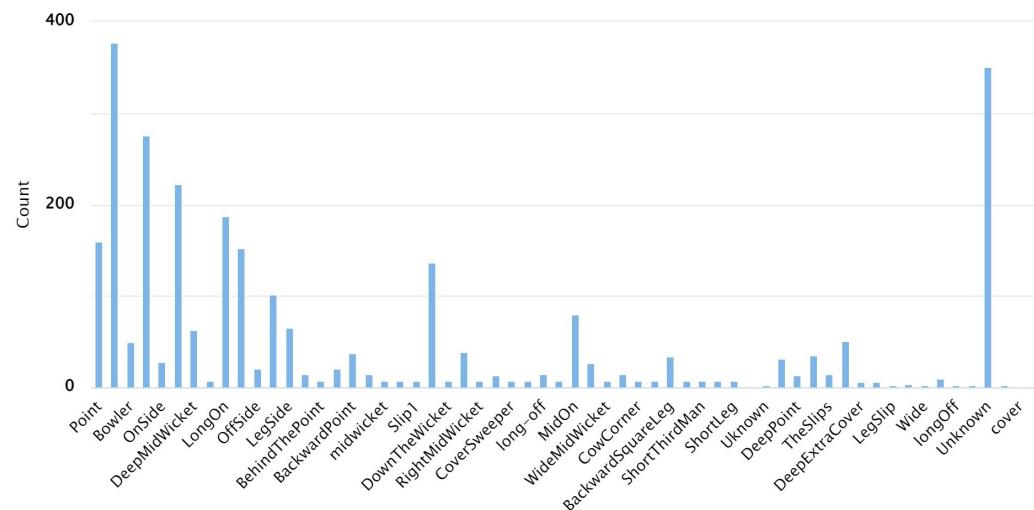


Figure 8. Distribution of Fielding Positions.

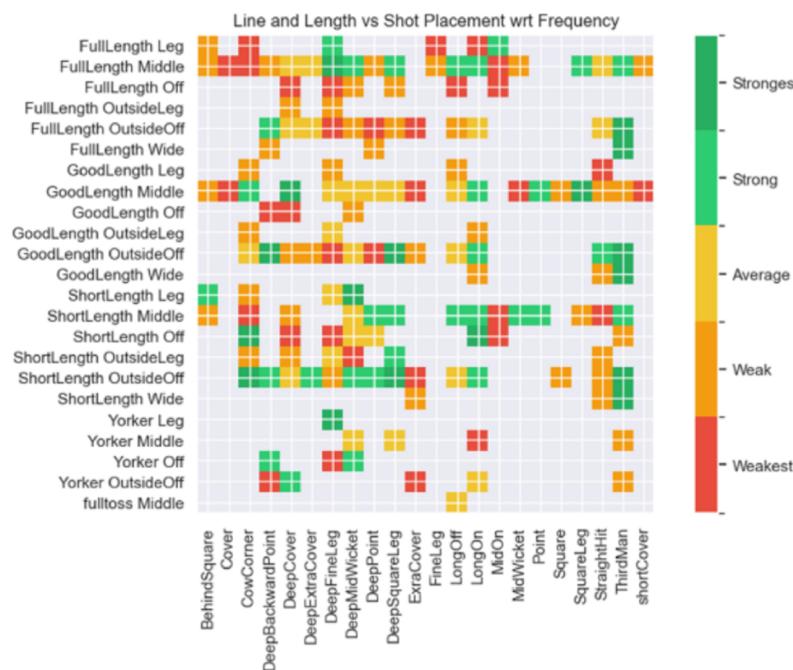


Figure 9. Strengths & Weaknesses: Delivery Line, Length vs. Scoring Zone (Frequency).

Table 3. Feature Combinations, Performance Index (PI), and Frequency Ratio (FR).

Delivery Length	Delivery Line	Field Position	PI	FR
FullLength	Leg	Behind square	1.41	0.0122
Goodlength	Middle	Extra cover	0.17	0.0031
shortlength	Off	Deep mid wicket	1.05	0.0072
Yorker	OutsideLeg	Backward square	0.00	0
FullToss	OutsideOff	ThirdMan	0.72	0.0061
FullLength	Wide	Gully	0.64	0.0083
GoodLength	Middle	Deep Gully	0.53	0.0094
FullLength	Off	Point	0.16	0.0117
ShortLength	OutsideOff	LongOff	1.08	0.0182
Yorker	Wide	Longon	0.51	0.0032
FullToss	Off	DeepFineLeg	1.52	0.0142

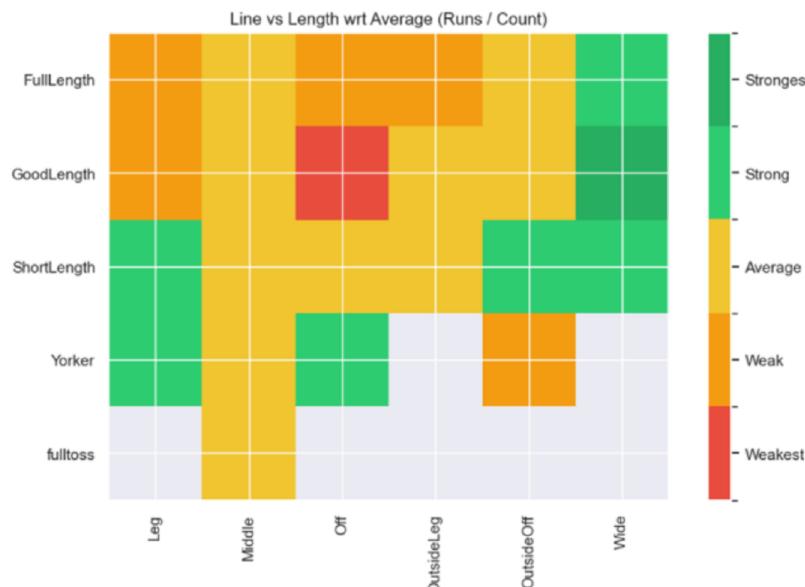


Figure 10. Strong and Weak Line & Length Based on Average Runs.

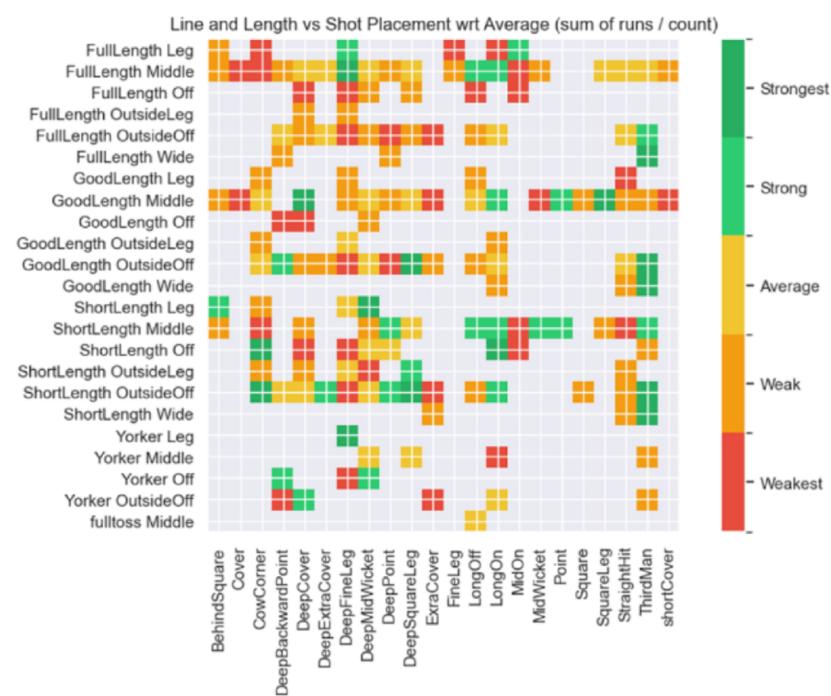


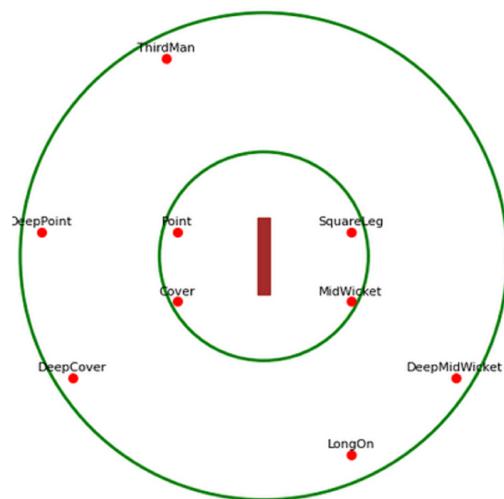
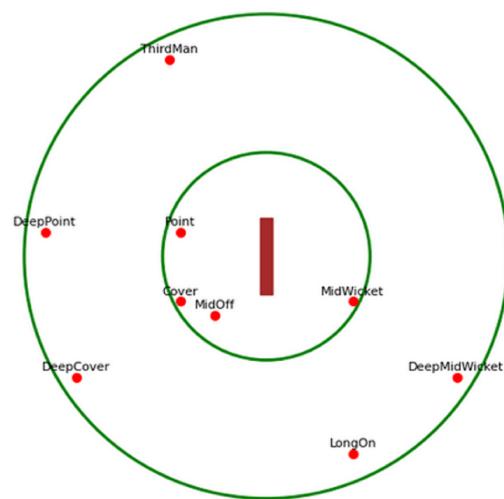
Figure 11. Strengths and Weaknesses: Delivery Line & Length vs. Field Positions.

4.5. Strategy Planner—Crafting Optimal Bowling and Fielding Strategies

In the context of analyzing the talented batsman under study, the planning step is pivotal. It transforms the insights derived from the analyzer module into precise actionable cricket strategies. Figures 12–15 showcase the field placement strategies based on line and length combinations, highlighting the precision of the planner step in providing a competitive edge by strategically restricting the opponent's scoring and disrupting the batter's rhythm. Table 4 provides a sample list of strategies, thus directly addressing the objectives outlined in Research Question 3 and identifying context-aware fielding and bowling strategies based on individual player attributes. The strategies are categorized considering bowler line, length, and specific match contexts such as powerplay, overs 7–15, and the last five overs.

Table 4. Field Placement Strategies.

Contexts (Overs, Line, Length)	Field Placement Strategy
Overs: 1–6 (Powerplay) Line: Middle Length: Good	mid-on, mid-off, mid-wicket, short fine-leg, cover, point, deep-mid-wicket, extra-cover, third-man, deep square-leg
Overs: 7–15 (Middle overs) Line: Outside Leg Length: Fuller	mid-off, mid-wicket, square-leg, short third-man, point, deep fine-leg, long-on, deep-mid-wicket, long leg
Overs: 7–15 (Middle overs) Line: Outside Off Length: Good	long-off, deep mid-wicket, deep extra-cover, third-man, cow corner, point, cover, fine-leg, square-leg
Overs: 15–20 (Last overs) Line: Middle Length: Yorker	mid-on, mid-wicket, square-leg, cover, point, third-man, long-on, deep-mid-wicket, extra-cover

**Figure 12.** Recommended Field Placement Strategy for Good Length and Outside Off Line Deliveries.**Figure 13.** Recommended Field Placement Strategy for Off Line Deliveries.

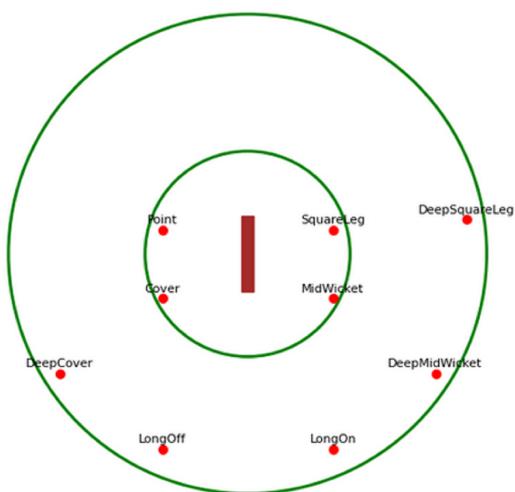


Figure 14. Recommended Field Placement Strategy for Full and Middle Line Deliveries.

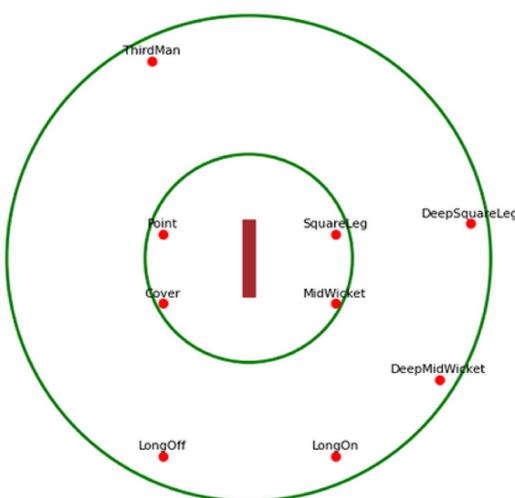


Figure 15. Recommended Field Placement Strategy for Yorker Length and Outside Leg Line Deliveries.

To evaluate the effectiveness of field placement strategies tailored to the performance of the prominent batsman against specific lengths and lines of delivery, simulations were conducted using data from the actual innings played by the batsman. The simulations aimed to assess how the suggested field placements would impact the runs scored by the selected player under study. The simulation process involved setting the field placement, as recommended by the framework, which was determined based on the delivery length and line. During the simulation, the following scenarios were considered.

- If the shot found a fielder inside the 30-yard circle, recorded 0 runs.
- If the shot found a fielder outside the 30-yard circle, recorded 1 run.
- If the shot didn't find any fielder, recorded 4 runs.

When simulating the targeted batsman's innings using the framework's bowling and fielding strategies, significant savings in the runs were observed. These strategies were applied to three main innings, as illustrated in Table 5. Figure 16 compares actual versus simulated cricket scores across three innings, showing significant run savings of 35.1%, 37.7%, and 26.4% respectively. After conducting a paired *t*-test to compare the actual scores with the simulated scores across the innings, we obtained a *t*-statistic of 7.20 and a *p*-value of 0.0187, as depicted in Figure 17. A low *p*-value (less than 0.05) indicates that the differences between the actual and simulated scores are statistically significant. The RunsGuard Framework strategy for optimizing field placements leads

to a significant reduction in runs scored against. This supports our hypothesis that the framework effectively saves runs compared with actual game outcomes. These results demonstrate the practical effectiveness of the research in optimizing field placements to restrict focused players' scoring opportunities. By applying these data-driven strategies, teams can potentially achieve substantial improvements in their ability to contain runs and effectively challenge prominent batters.

Table 5. Runs Saved.

Innings	Actual Score	Simulation Score	Saved
[46]	97	63	35.1%
[47]	122	76	37.7%
[48]	110	81	26.4%

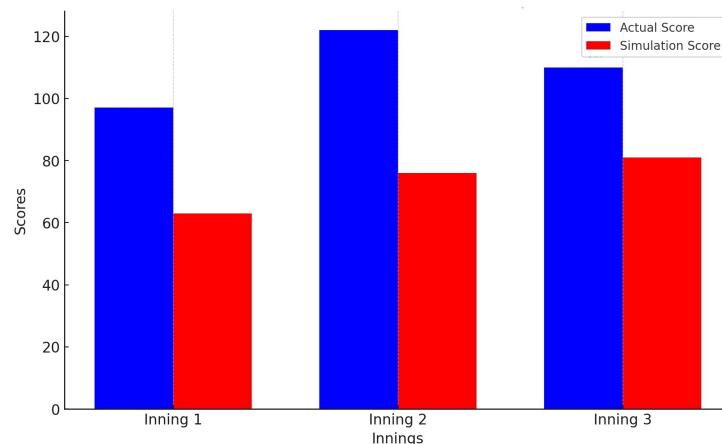


Figure 16. Actual vs. Simulation Results Comparison.

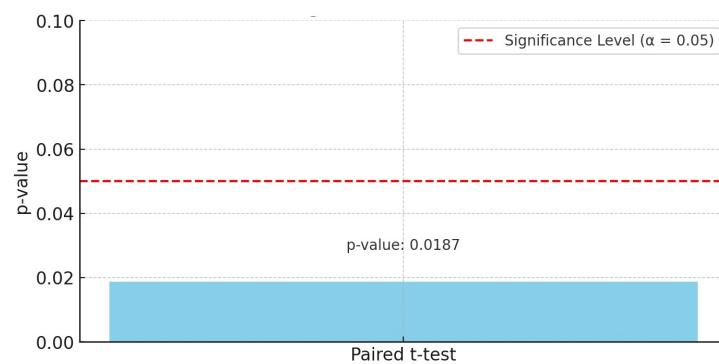


Figure 17. Statistical Significance of Runs Saved.

5. Limitations and Future Directions

In cricket analytics, the RunsGuard Framework provides a fundamental structure that establishes the foundations for future research. Developing an understanding of and improving the cricket strategy and analysis begins with this framework. Its adaptability and flexibility offer a solid basis for broadening the focus of the analysis, adding more profound insights, considering a broader range of contextual factors, and enhancing planning strategies. Our study mainly focuses on reducing runs as the main goal. But, this overlooks other important parts of cricket strategy. These include how well players take wickets and how well the team works together. Although this framework offers significant insights,

we recognize certain limitations inherent in our methodology, which open avenues for further refinement.

Feature Set Expansion: Adding factors, such as pitch conditions, player form, and opposition strategies, would have strengthened our analysis. With a broader set of features, these additions can significantly enhance the framework's predictive power and strategic support.

Adaptability Across Cricket Formats: Each cricket format required a tailored strategy. The general approach of the framework might not capture the specific demands of T20Is, T10s, Tests, and ODIs without incorporating format-specific analyses and adapting to dynamic environmental factors. Therefore, it is important to improve the proposed approach. It is necessary to focus on the unique challenges of each cricket format to provide precise and relevant strategies for all types of cricket games.

Player Performance Variability: Players often learn, practice, and change how they play. If we do not sufficiently focus on their recent approaches and simply use all data in the same way, we might not come up with the best strategies. In the future, we need to examine how giving more weight to recent performance can improve our framework.

Consideration of Physical Conditions: Currently, our model does not consider how players' physical and mental health impacts the strategy. Adding these factors to the model could deepen the analysis.

Psychological Factors: Our model does not consider how game pressure affects the players. This hints at a complex link between mindset and performance, a topic that is ripe for future studies.

Data Source Biases and Accuracy: Our reliance on textual comments, in turn, raises the possibility of an inherent bias because they are exhaustive. Furthermore, there is a possibility that the feature extraction procedure is flawed owing to the use of natural language processing (NLP) approaches. These factors serve as the most important reminders of the ongoing need for methodological advancement.

Acknowledging these limitations not only sets the stage for future research but also underscores our commitment to advancing the field of sports analytics. Each limitation represents a stepping stone towards more nuanced, comprehensive models that more closely mirror the complexities of cricket.

6. Conclusions

This study introduced a framework involving four key steps: data acquisition, information extraction, analytical insight generation, and strategic planning. The architecture uses analytical methods and empirical data to optimize cricket bowling and fielding tactics. This fits well with the current tendency toward sports decision-making based on statistics. A case study centered on a prominent cricketer was conducted to practically demonstrate the framework's efficacy. In-depth analysis, dataset extraction from textual match commentary, and dataset collection from reliable sources were all included in the present research. The findings of this case study are convincing. The application of bowling and fielding tactics in simulated innings for the subject athlete consistently led to significant reductions in the number of runs scored. There was a noteworthy 35.1% run decrease in the first inning, and even more astounding 37.7% and 26.4% run savings in the second and third innings, respectively. These results highlight how the findings may be used to optimize field placements and reduce scoring possibilities for star batsmen, which is the subject of this study. Coaches, captains, and players in cricket teams can gain a competitive advantage by using this paradigm, which provides data-driven insights. This research serves as a bridge between data analytics and cricket strategy, enabling teams to elevate their performance and competitiveness significantly. The general methodology of the RunsGuard Framework provides a blueprint for data-driven strategic decision making. This can be customized to meet the needs of various sports. Baseball, like a cricket, is a game in which player positioning and analysis of player actions play a crucial role in

influencing the game's outcome. In the context of baseball, the framework can be used to enhance the strategic elements of the game by adapting the specifics of the data collection and analysis. This shift towards a more analytical and evidence-backed way of thinking helps not only in playing the game smarter but also in managing sports with a keen eye on solid facts and figures. Future research should delve into dynamic analysis to adapt field placements and bowling strategies in real time, incorporating environmental factors such as pitch, weather, and dew conditions to enhance accuracy. Additionally, research may extend to developing a contextual analysis that considers additional factors, including the required run rate, available bowling resources, and the number of outs for context-aware recommendations. This extension aims to align individual strengths with team strategies and create real-time decision-support systems to enhance in-game strategies.

Author Contributions: The primary research work was conducted by A.H. (Aatif Hussain) under the supervision of S.A. and A.H. (Awais Hassan). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this research paper:

AI	Artificial intelligence
ICC	International Cricket Council
IPL	Indian Premier League's
NLP	National Language Processing
ML	Machine Learning
ANN	Artificial Neural Network
RF	Random Forest
KNN	K-nearest Neighbors
SVM	Support Vector Machine
ODI	One-Day International
LR	Logistic Regression
NB	Naive Bayes
CIR	Commitment Index Rank
CRF	Conditional Random Fields
PI	Performance Index
FR	Frequency Ratio
DOM	Document Object Model
PCA	Principal Component Analysis
T20I	Twenty Twenty International
T10	Ten Ten
NLTK	Natural Language Toolkit

References

1. Avci, P.; Bayrakdar, A. Revolutionizing Sport—How Technology is Changing the Sports Industry? In *The Use of Developing Technology in Sports*; Ozgur Press: Gaziantep, Turkey, 2023. [CrossRef]
2. Sports Technology Market—Global Size, Share & Industry Analysis. Available online: <https://www.marketsandmarkets.com/Market-Reports/sports-technology-market-104958738.html> (accessed on 9 October 2023).
3. Vlinkinfo. How Data Analytics is Being Utilized in The Sports Industry? Available online: <https://www.vlinkinfo.com/blog/data-analytics-in-sports-industry/> (accessed on 12 October 2023).

4. TSS-Admin. Importance of Technology. *The Sports School*, 5 2023. Available online: <https://thesportsschool.com/importance-of-technology-in-sports> (accessed on 12 October 2023).
5. Biswas, M.; Niamat Ullah Akhund, T.M.; Mahbub, M.K.; Saiful Islam, S.M.; Sorna, S.; Shamim Kaiser, M. A survey on predicting player's performance and team recommendation in game of cricket using machine learning. In Proceedings of the Information and Communication Technology for Competitive Strategies (ICTCS 2020) ICT: Applications and Social Interfaces, Jaipur, India, 11–12 December 2020; Springer: Singapore, 2022; pp. 223–230.
6. ICC. Annual Report, 2023. Available online: <https://www.icc-cricket.com/about/the-icc/publications/annual-report> (accessed on 3 January 2024).
7. Bucaille, A.; Jarvis, D.; Lee, P.; Giorgio, P.; Westcott, K. Live Sports: The Next Arena for the Streaming Wars | Deloitte. 2022. Available online: <https://www2.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/2023/live-sports-streaming-wars.html> (accessed on 7 November 2023).
8. Mishra, U. How Is Technology Used in Cricket? | Analytics Steps. Available online: <https://analyticssteps.com/blogs/how-technology-used-cricket> (accessed on 10 November 2023).
9. Ariyaratne, M.K.A.; Silva, R.M. Meta-heuristics meet sports: A systematic review from the viewpoint of nature-inspired algorithms. *Int. J. Comput. Sci. Sport* **2022**, *21*, 49–92. [CrossRef]
10. Lim, J.; Wong, S.; McErlain-Naylor, S.A.; Scanlan, A.; Goggins, L.; Ahmun, R.; Comfort, P.; Weldon, A. Strength and conditioning for cricket fielding: A narrative review. *Strength Cond. J.* **2023**, *45*, 509–524. [CrossRef]
11. Behera, S.R.; Agrawal, P.; Awekar, A.; Vedula, V.S. Mining strengths and weaknesses of cricket players using short text commentary. In Proceedings of the 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 673–679.
12. Miraoui, Y. Analyzing sports commentary in order to automatically recognize events and extract insights. *arXiv* **2023**, arXiv:2307.10303.
13. Arif, S.; Umair, M.; Naqvi, S.M.K.; Ikram, A.; Ikram, A. Detection of bowler's strong and weak area in cricket through commentary. In Proceedings of the 2nd International Conference on Future Networks and Distributed Systems, Amman, Jordan, 26–27 June 2018; pp. 1–14.
14. Roy, R.; Rahman, M.R.; Shamim Kaiser, M.; Arefin, M.S. Developing a Text Mining Framework to Analyze Cricket Match Commentary to Select Best Players. In Proceedings of the International Conference on Big Data, IoT, and Machine Learning: BIM 2021, Cox's Bazar, Bangladesh, 23–25 September 2021; Springer: Singapore, 2022; pp. 217–229.
15. GOEL, A.; Singh, P.P. Impact Calculation of the Players Using the Cricket Commentary Corpus. *Eng. Open Access* **2023**, *1*, 48–53.
16. Balasundaram, A.; Kothandaraman, D.; Prashanth, B.; Ashokkumar, S. Predicting different facets in the game of cricket using machine learning. *AIP Conf. Proc.* **2022**, *2418*, 020027.
17. Behera, S.R.; Saradhi, V.V. Learning strength and weakness rules of cricket players using association rule mining. In Proceedings of the International Workshop on Machine Learning and Data Mining for Sports Analytics, Virtual, 13 September 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 79–92.
18. Rauf, M.A.; Ahmad, H.; Faisal, C.N.; Ahmad, S.; Habib, M.A. Extraction of Strong and Weak Regions of Cricket Batsman through Text-Commentary Analysis. In Proceedings of the 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 5–7 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
19. Shetty, M.; Rane, S.; Pandita, C.; Salvi, S. Machine learning-based Selection of Optimal sports Team based on the Players Performance. In Proceedings of the 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 10–12 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1267–1272.
20. Puram, P.; Roy, S.; Srivastav, D.; Gurumurthy, A. Understanding the effect of contextual factors and decision making on team performance in Twenty20 cricket: An interpretable machine learning approach. *Ann. Oper. Res.* **2023**, *325*, 261–288. [CrossRef]
21. Kapadia, K.; Abdel-Jaber, H.; Thabtah, F.; Hadi, W. Sport analytics for cricket game results using machine learning: An experimental study. *Appl. Comput. Inform.* **2020**, *18*, 256–266. [CrossRef]
22. Kumar, C.; Balasubramanian, G. Comparative analysis of pitch ratings in all formats of cricket. *Manag. Labour Stud.* **2023**, *48*, 0258042X221148069. [CrossRef]
23. Das, N.R.; Ghosh, S.; Mukherjee, I.; Paul, G. Adoption of a ranking based indexing method for the cricket teams. *Expert Syst. Appl.* **2023**, *213*, 118796. [CrossRef]
24. Das, D.; Saikia, H.; Bhattacharjee, D. Optimal playing strategies of a batsman against bowling type in limited-over cricket: An application of game theory. *Commun. Stat. Case Stud. Data Anal. Appl.* **2022**, *8*, 738–751. [CrossRef]
25. Zia, S.; Liaqat, H.B.; Zia, H.U.; Kong, X.; Shamshad, S. Player-aware resource compensation in interrupted cricket matches. *PeerJ Comput. Sci.* **2022**, *8*, e917. [CrossRef]
26. Nicholls, S.; Pote, L.; Thomson, E.; Theis, N. The Change in Test Cricket Performance Following the Introduction of T20 Cricket: Implications for Tactical Strategy. *Sports Innov. J.* **2023**, *4*, 1–16. [CrossRef]
27. Das, D.; Saikia, H.; Bhattacharjee, D.; Kushvaha, B. On estimating shot selection by a batsman in Twenty20 cricket: A probabilistic approach. *Commun. Stat. Case Stud. Data Anal. Appl.* **2022**, *8*, 354–367. [CrossRef]
28. Mody, K.; Malathi, D.; Jayaseeli, J.D. An Artificial Neural Network Approach for Classifying Cricket Batsman's Performance by Adam Optimizer and Prediction by Derived Attributes. In Proceedings of the Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 9–10 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–7.

29. Tirtho, D.; Rahman, S.; Mahbub, M.S. Cricketer's tournament-wise performance prediction and squad selection using machine learning and multi-objective optimization. *Appl. Soft Comput.* **2022**, *129*, 109526. [[CrossRef](#)]
30. Mahbub, M.K.; Miah, M.A.M.; Islam, S.M.S.; Sorna, S.; Hossain, S.; Biswas, M. Best Eleven Forecast for Bangladesh Cricket Team with Machine Learning Techniques. In Proceedings of the 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Dhaka, Bangladesh, 18–20 November 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
31. Pussella, P.; Silva, R.M.; Egodawatta, C. In-game winner prediction and winning strategy generation in cricket: A machine learning approach. *Int. J. Sports Sci. Coach.* **2022**, *18*, 17479541221119738. [[CrossRef](#)]
32. Awan, M.J.; Gilani, S.A.H.; Ramzan, H.; Nobanee, H.; Yasin, A.; Zain, A.M.; Javed, R. Cricket match analytics using the big data approach. *Electronics* **2021**, *10*, 2350. [[CrossRef](#)]
33. Sarangi, S.; Singh, R.R. Winning one-day international cricket matches: A cross-team perspective. *J. Bus. Anal.* **2023**, *6*, 39–58. [[CrossRef](#)]
34. Sajib, A.H.; Limon, S.H.; Naser, A.I.; Saha, G. Analysing factors impacting Bangladesh men's T20 cricket performance. *Sci. J. Sport Perform.* **2023**, *3*, 79–94. [[CrossRef](#)]
35. Pramanik, M.A.; Hasan Suzan, M.M.; Biswas, A.A.; Rahman, M.Z.; Kalaiarasi, A. Performance Analysis of Classification Algorithms for Outcome Prediction of T20 Cricket Tournament Matches. In Proceedings of the 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 25–27 January 2022; pp. 1–7. [[CrossRef](#)]
36. Krishnan, S.; Vasan, R.V.; Varma, P.; Mala, T. Probabilistic Forecasting of the Winning IPL Team Using Supervised Machine Learning. In Proceedings of the Advanced Network Technologies and Intelligent Computing, Varanasi, India, 22–24 December 2022; Woungang, I., Dhurandher, S.K., Pattanaik, K.K., Verma, A., Verma, P., Eds.; Springer: Cham, Switzerland, 2023; pp. 138–152.
37. Yang, Y.; Zhang, B.; Guo, D.; Xu, R.; Wang, W.; Xiong, Z.; Niyato, D. Semantic Sensing Performance Analysis: Assessing Keyword Coverage in Text Data. *IEEE Trans. Veh. Technol.* **2023**, *72*, 15133–15137. [[CrossRef](#)]
38. Yang, Y.; Zhang, B.; Guo, D.; Xu, R.; Kumar, N.; Wang, W. Mean Field Game and Broadcast Encryption-Based Joint Data Freshness Optimization and Privacy Preservation for Mobile Crowdsensing. *IEEE Trans. Veh. Technol.* **2023**, *72*, 14860–14874. [[CrossRef](#)]
39. Rahimian, P.; Toka, L. A data-driven approach to assist offensive and defensive players in optimal decision making. *Int. J. Sports Sci. Coach.* **2023**, *19*, 245–256. [[CrossRef](#)]
40. Jha, A.; Kar, A.K.; Gupta, A. Optimization of team selection in fantasy cricket: A hybrid approach using recursive feature elimination and genetic algorithm. *Ann. Oper. Res.* **2023**, *325*, 289–317. [[CrossRef](#)]
41. Anuraj, A.; Boparai, G.S.; Leung, C.K.; Madill, E.W.R.; Pandhi, D.A.; Patel, A.D.; Vyas, R.K. Sports Data Mining for Cricket Match Prediction. In Proceedings of the Advanced Information Networking and Applications, Juiz de Fora, Brazil, 29–31 March 2023; Barolli, L., Ed.; Springer: Cham, Switzerland, 2023; pp. 668–680.
42. Raja Subramanian, R.; Vijaya Karthick, P.; Dhanasekaran, S.; Raja Sudharsan, R.; Hariharasitaraman, S.; Rajasekaran, S.; Murugan, B.S. Regression to Forecast: An In-Play Outcome Prediction for One-Day Cricket Matches. In Proceedings of the International Conference on Cognitive and Intelligent Computing, Hyderabad, India, 11–12 December 2021; Kumar, A., Ghinea, G., Merugu, S., Hashimoto, T., Eds.; Springer: Singapore, 2023; pp. 43–55.
43. ESPNcricinfo. Today's Cricket Match | Cricket Update | Cricket News | ESPNcricinfo. Available online: <https://www.espncricinfo.com/> (accessed on 20 September 2023).
44. Cricbuzz. Live Cricket Score, Schedule, Latest News, Stats & Videos. Available online: <https://www.cricbuzz.com/> (accessed on 20 September 2023).
45. Limited, E.D.M.P. Babar Azam Profile—Cricket Player Pakistan | Stats, Records, Video. 2023. Available online: <https://www.espncricinfo.com/cricketers/babar-azam-348144> (accessed on 14 November 2023).
46. Limited, E.D.M.P. PAK vs. WI, WI in Pakistan 2018, 2nd T20I at Karachi, 2 April 2018—Full Scorecard. Available online: <https://www.espncricinfo.com/series/wi-in-pakistan-2018-1140067/pakistan-vs-west-indies-2nd-t20i-1140070/full-scorecard> (accessed on 20 November 2023).
47. Limited, E.D.M.P. SA vs. PAK, Pakistan Tour of South Africa 2021, 3rd T20I at Centurion, 14 April 2021—Full Scorecard. Available online: <https://www.espncricinfo.com/series/pakistan-tour-of-south-africa-2021-1251556/south-africa-vs-pakistan-3rd-t20i-1251577/full-scorecard> (accessed on 11 December 2023).
48. Limited, E.D.M.P. PAK vs. ENG, England in Pakistan 2022, 2nd T20I at Karachi, 22 September 2022—Full Scorecard. Available online: <https://www.espncricinfo.com/series/england-in-pakistan-2022-1327226/pakistan-vs-england-2nd-t20i-1327229/full-scorecard> (accessed on 11 December 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.