Crick-XI: A Performance-Based XI Selector

Aliyan Aliyan
Student of Artificial Intelligence and Data Engineering Department
Istanbul Technical University)
Istanbul, Turkey
aliyan23@itu.edu.tr

***Abstract*—Abstract:**

The Performance-Based XI Selector is a pioneering initiative poised to revolutionize cricket team selection by harnessing the power of data mining and Association Rule algorithms. This project embarks on a comprehensive analysis of match data, aiming to unearth statistically significant associations between player attributes—such as batting position, strike rate, and scoring ranges—and match outcomes. The primary objective is to furnish cricket teams with actionable insights that optimize player selections, ultimately elevating team performance to new heights.

The Crick-XI project employs a meticulous exploration of association rules, striving to make meaningful contributions to the decision-making processes within the cricketing community. By providing a data-driven approach to shaping winning combinations, Crick-XI seeks to bridge the gap between traditional cricket team selection methods and the evolving landscape of sports analytics. This initiative represents a significant leap toward more informed and strategic cricket team selections, acknowledging the increasingly pivotal role of data analytics in the realm of sports.

In addition to the identification of statistical associations, the project goes further by generating performance scores for each rule. These scores are derived through a meticulous combination of various performance metrics, ensuring a nuanced and comprehensive evaluation of the rules discovered through association rule mining. By incorporating these scores, the project aims to provide nuanced insights into the significance and impact of each identified rule on team performance.

Through its data-centric approach, the Performance-Based XI Selector aspires to contribute not only to the understanding of cricket dynamics but also to the practical implementation of these insights in the decision-making processes of cricket teams. This initiative stands as a testament to the growing synergy between data analytics and sports, offering a novel perspective on cricket team selection and strategic decision-making.

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Cricket, an enduring sport embraced by diverse cultures worldwide, is at the brink of a transformative era where age-old strategies meet the cutting-edge precision of data analytics. Amidst this dynamic landscape, the process of team selection, once reliant on intuition and experience, is undergoing a paradigm shift. This project, aptly named "Crick-XI: A Performance-Based XI Selector," signifies a pioneering venture into the realms of cricket analytics, employing advanced data mining and Association Rule algorithms to redefine how cricket teams are composed and strategized.

The core objective of Crick-XI is to unravel the intricate tapestry of relationships between individual player attributes and match outcomes. Harnessing the power of data mining, particularly through Association Rule Mining (ARM), the project seeks to unearth statistically significant patterns within vast datasets. Through this, Crick-XI aims to equip cricket teams with actionable insights, facilitating nuanced player selections and, consequently, elevating overall team performance.

In an era where the fusion of sports and data science is reshaping traditional paradigms, Crick-XI emerges as a pivotal initiative. Traditional approaches to team selection are being augmented by a data-driven methodology, acknowledging that strategic decisions based on statistical insights enhance a team's competitive edge. This project addresses the pressing need for cricket teams to adapt to this evolving landscape, where success is increasingly intertwined with the ability to harness data for strategic advantage.

To comprehend the intricacies explored in this project, a foundational understanding of cricket, its nuances, and the specifics of the T20 format is imperative. Cricket, a sport contested between two teams of 11 players each, involves the team batting to accrue runs while the bowling team endeavors to dismiss batsmen and curtail the opponent's score. The T20 format, distinguished by its rapid pace, introduces distinct challenges and opportunities, making it a focal point for our research.

Inspired by the groundbreaking studies that precede it, this project draws from the insights garnered in "Association Rule Mining Approach in Strategy Planning for Team India in ICC World Cup 2015." The application of Association Rule Mining in cricket analytics has proven instrumental in deciphering player intricacies and the multifaceted factors influencing team success. Additionally, the research titled "Selection for Balanced Cricket Team Fourthcoming ICC Championship 2017" serves as a cornerstone, extending association rule mining techniques to the nuanced art of crafting a balanced cricket team.

As we embark on a comprehensive exploration of Crick-XI, the subsequent sections will unfurl the intricacies of the methodology, delve into the nuances of data analysis, and present the project's findings, offering a holistic perspective on its implications for the future of cricket team selection.

II. Related Work

In exploring the landscape of cricket analytics, prior studies have paved the way for understanding player performance

and team strategy. Two key works in association rule mining (ARM) provide valuable insights, influencing the current project's objectives:

A. "Association Rule Mining Approach in Strategy Planning for Team India in ICC World Cup 2015"

This research, centered around the ICC World Cup 2015, employed ARM techniques to uncover hidden relations within individual Indian players' career records. The study aimed to reveal intrinsic factors impacting player performances, including the venue of the match, batting position, strike rate, and runs. It introduced scoring for each rule, offering a quantitative measure of significance. While the study contributed valuable insights, it acknowledged limitations related to the quality of opposing teams. The present project builds upon this work, emphasizing the importance of quantifying the significance of association rules in shaping team strategy.

B. "Selection for Balanced Cricket Team Fourthcoming ICC Championship 2017"

Anticipating the ICC Championship 2017, this study explored the use of ARM in constructing a balanced cricket team. The analysis encompassed a comprehensive dataset containing player statistics, such as grounds, venue, strike rate, average, and various performance metrics for batsmen and bowlers. The study introduced scoring for each rule to prioritize key associations. It acknowledged challenges in team selection, including environmental and player-specific factors. The current project draws inspiration from this work, incorporating a scoring mechanism to assess the significance of association rules in the context of team selection.

C. Identified Gaps and Novelty

While these existing studies contribute significantly to cricket analytics, they reveal certain gaps and limitations:

1) Holistic Player Performance Metrics with Scoring: In contrast to the previous studies' focus on specific player attributes, the current project introduces a holistic approach by considering a broader set of player performance metrics. The introduction of scoring for each rule adds a quantitative dimension, allowing for the assessment of the significance of each association rule. This approach contributes to a more nuanced and multifaceted evaluation of player contributions and team strategies.

The identified gaps and the introduction of a scoring mechanism underscore the novelty and significance of the current project. By quantifying the significance of association rules and addressing existing limitations, the project aims to advance the field of cricket analytics, providing actionable insights for optimizing player selections and enhancing team performance.

III. Proposed Work

The Crick-XI initiative strives to advance the field of cricket analytics, building upon the groundwork laid by prior studies[1][2]. Our primary focus revolves around revolutionizing cricket team selection through the application of Association Rule Mining (ARM) techniques. This section provides a detailed overview of the proposed work, encompassing data collection, preprocessing, association rule mining, rule evaluation, implementation strategies, hypotheses, expected outcomes, and the theoretical framework.

A. Methodology

The proposed work employs a robust methodology centered around Association Rule Mining (ARM) to extract meaningful insights from cricket match data. The key steps of the methodology include:

1) Data Collection: Initially intending to use the T20 World Cup 2021 dataset as outlined in the project proposal, a critical reassessment led to the realization that its scope and duration might not adequately capture the requisite diversity and dynamics for a thorough analysis. In response, an alternative source was sought, resulting in the acquisition of a comprehensive ball-by-ball dataset covering all T20 matches since the format's inception from Kaggle. This expanded dataset offers a broader temporal and contextual spectrum, facilitating a more in-depth exploration of patterns and trends in cricket player performance.

The shift in dataset selection aligns with the project's commitment to ensuring robustness, depth, and inclusivity in the data analysis process, enhancing the richness of the study for a more nuanced exploration of association rules.

2) Preprocessing: Data preprocessing involves cleaning and organizing the dataset to ensure its suitability for association rule mining. This includes handling missing values, standardizing formats, and addressing any outliers.

3) Association Rule Mining: Data preprocessing plays a crucial role in ensuring the dataset's readiness for analysis. This involves addressing missing values, handling outliers, and ensuring uniformity in data representation. Feature engineering techniques are employed to extract relevant attributes such as batting position, strike rate, scoring ranges, and opponent teams.

4) Scoring Mechanism: To assess the significance of association rules, a scoring mechanism has been introduced. The scoring is based on a combination of performance metrics, allowing for a nuanced evaluation of the impact of each rule.

B. Tools and Techniques

The project leverages the Python programming language along with popular data science libraries, including Pandas for data manipulation and Scikit-learn for implementing the Apriori algorithm. The heart of the project lies in the application of Association Rule Mining techniques, with a specific focus on the Apriori algorithm. This algorithm is instrumental in unveiling hidden patterns and relationships within the dataset[1][2]. The analysis explores factors influencing player performance, including venue, batting order, and match conditions, to derive association rules that will inform team selection.

C. Rule Evaluation

The quality and significance of association rules are assessed based on parameters such as support, confidence, lift,

Jaccard coefficient, and Kulczynski measure. Through iterative refinement, rules are enhanced to improve their relevance and effectiveness in informing team selection strategies.

D. Hypotheses and Expected Outcomes

Hypothesis 1: Significant associations exist between specific player attributes and match outcomes. **Hypothesis 2:** Association rules derived from the dataset provide actionable insights for cricket team selection, leading to improved team performance.

E. Theoretical Framework

The project's theoretical framework draws upon the principles of Association Rule Mining, emphasizing the identification of statistically significant relationships among diverse player attributes and their impact on match outcomes. The Apriori algorithm, a cornerstone of this framework, operates by identifying frequent itemsets and generating association rules, establishing a solid foundation for uncovering hidden patterns in cricket match data.

F. Expected Outcomes

The project anticipates several key outcomes:

1) Identification of Significant Association Rules: Through association rule mining and the introduced scoring mechanism, the project aims to identify and prioritize significant rules that correlate with player performance and match outcomes.

2) Improved Team Selection Strategies: The insights gained from the analysis are expected to contribute to more informed team selection strategies. By understanding the associations between various player attributes and success on the field, the project seeks to optimize team compositions.

3) Enhanced Decision-Making for Team Management: The project aspires to provide actionable insights for team management, enabling better decision-making in areas such as strategy planning, batting order adjustments, and match-specific player roles.

IV. Expreimental Results

Following are the association rules Generated through apriori algorithm , along with there confidence and lift, score table represent the significance of each rule by giving the score (0 to 1), score is the combination of 4 evaluation metrics, (0.30 X confidence, 0.10 X lift, 0.30 X jaccard's coeffient and 0.30 X kulczynski measure). The weights were for each matric were decided by testing different values, and my determining how much weights on each metrics as making change to score value, lift values have a very huge range compare to other attributes, thats why it is alloted very less wait, score was calculated with the normalized values of each atribute, min max normalization was use to normalized the values. Table I and II refers to batsman and III and IV refers to bowler

TABLE I
GENERATED ASSOCIATION RULES FOR BABAR AZAM (WON)

Antecedents	Consequents	Confidence	Lift	Score
Runs >= 50, Babar Azam, New Zealand, S/R >= 120, Top Order	Won	1.000	1.738	0.604
Runs >= 50, Babar Azam, West Indies, S/R >= 120, Top Order	Won	1.000	1.738	0.604
2nd Inning, Babar Azam, Runs >= 50, Top Order	Won	0.929	1.614	0.557
Babar Azam, West Indies	Won	0.929	1.614	0.557
2nd Inning, Babar Azam, Runs >= 50	Won	0.929	1.614	0.557
2nd Inning, Babar Azam, Runs >= 50, S/R >= 120	Won	0.923	1.605	0.550
Bangladesh, Babar Azam, Runs <25	S/R <90, Won	1.000	6.551	0.533
Babar Azam, West Indies	Top Order, Won	0.929	3.620	0.507
2nd Inning, Babar Azam, Runs >= 50	Top Order, Won	0.929	3.620	0.507
2nd Inning, Babar Azam, Runs >= 50, S/R >= 120	Top Order, Won	0.923	3.599	0.497
Babar Azam, 1st Inning, West Indies	Won	0.875	1.521	0.496
Babar Azam	Top Order, Won	0.670	2.613	0.482
Babar Azam, S/R >= 90 and <120, West Indies	Won	0.857	1.490	0.477
2nd Inning, Babar Azam, Runs >= 50	S/R >= 120, Top Order, Won	0.857	6.711	0.471

TABLE II
GENERATED ASSOCIATION RULES FOR BABAR AZAM (LOST)

Antecedents	Consequents	Confidence	Lift	Score
Babar Azam, S/R >= 90 and <120, 1st Inning, Runs >= 25 and <50	Lost	0.667	1.570	0.295
Babar Azam, England	Top Order, Lost	0.600	4.187	0.292
Babar Azam, Top Order, England	Lost	0.600	1.413	0.251
Babar Azam, England	Lost	0.600	1.413	0.251
Sri Lanka, Babar Azam	Lost	0.625	1.472	0.248
Sri Lanka, Babar Azam, Top Order	Lost	0.625	1.472	0.248
Sri Lanka, Babar Azam	Top Order, Lost	0.625	4.361	0.241
Babar Azam, S/R >= 120, England	Top Order, Lost	0.600	4.187	0.235
Babar Azam, 1st Inning, England	Top Order, Lost	0.600	4.187	0.235
Babar Azam, England, S/R >= 120	Lost	0.600	1.413	0.231
Babar Azam, 1st Inning, England	Lost	0.600	1.413	0.231
Babar Azam, 1st Inning, Top Order, England	Lost	0.600	1.413	0.231
Babar Azam, Top Order, S/R >= 120, England	Lost	0.600	1.413	0.231
Babar Azam, 1st Inning, Runs >= 25 and <50	Lost	0.583	1.373	0.222
Babar Azam, 1st Inning, Runs >= 25 and <50	S/R >= 90 and <120, Lost	0.500	5.534	0.196
Babar Azam, S/R >= 90 and <120, 1st Inning	Lost	0.526	1.239	0.186
Babar Azam, S/R >= 90 and <120, Runs >= 25 and <50	Lost	0.500	1.177	0.141
Babar Azam, 1st Inning, Runs >= 25 and <50, Top Order	Lost	0.500	1.177	0.128

TABLE III
GENERATED ASSOCIATION RULES FOR SHADAB KHAN (WON)

Antecedents	Consequents	Confidence	Lift	Score
Shadab Khan, Over: 4, West Indies	Won	1.000	1.551	0.629
Shadab Khan, West Indies, Economy: 0-6	Won	1.000	1.551	0.625
Shadab Khan, Bangladesh	Won	1.000	1.551	0.620
Shadab Khan, Over: 4, West Indies, Economy: 0-6	Won	1.000	1.551	0.620
Shadab Khan, Wickets: 2-3, Economy: 6-8	Won	1.000	1.551	0.620
Shadab Khan, 1st Inning, Economy: 0-6	Won	0.917	1.421	0.564
Shadab Khan, 2nd Inning, West Indies, Over: 4	Economy: 0-6, Won	1.000	3.973	0.559
Shadab Khan, Bangladesh, Economy: 6-8	Over: 4, Won	1.000	2.509	0.558
Shadab Khan, Wickets: 2-3, 1st Inning, Economy: 6-8	Over: 4, Won	1.000	2.509	0.558
Shadab Khan, Wickets: 2-3, West Indies, Economy: 0-6	Over: 4, Won	1.000	2.509	0.558
Shadab Khan, West Indies, 1st Inning	Over: 4, Won	1.000	2.509	0.558
Shadab Khan, West Indies	Won	0.909	1.410	0.552
Shadab Khan, Wickets: 2-3, Over: 4, Economy: 0-6	Won	0.909	1.410	0.552
Shadab Khan, Over: 4, 1st Inning, Economy: 0-6	Won	0.909	1.410	0.552
Shadab Khan, Wickets: 2-3, Economy: 0-6	Won	0.909	1.410	0.552
Shadab Khan, Economy: 0-6, Wickets: 3+	2nd Inning, Won	1.000	3.050	0.550
Shadab Khan, Wickets: 2-3	Won	0.846	1.312	0.549
Shadab Khan, Over: 4, Wickets: 2-3	Won	0.840	1.303	0.539
Shadab Khan	Won	0.671	1.040	0.534
Shadab Khan, Wickets: 2-3, Economy: 0-6	Over: 4, Won	0.909	2.280	0.519
Shadab Khan, Over: 4	Won	0.703	1.090	0.516
Shadab Khan, 2nd Inning, Wickets: 2-3	Won	0.857	1.329	0.515
Shadab Khan, Wickets: 2-3	Over: 4, Won	0.808	2.026	0.505

TABLE IV
GENERATED ASSOCIATION RULES FOR SHADAB KHAN (LOST)

Antecedents	Consequents	Confidence	Lift	Score
Shadab Khan, Over: 4, 1st Inning, Sri Lanka	Lost	0.800	2.253	0.411
Shadab Khan, Over: 4, Sri Lanka	Lost	0.800	2.253	0.411
Shadab Khan, Over: 3, Economy: 10+	Lost	0.800	2.253	0.411
Wickets: 0-1, Shadab Khan, Over: 3, Economy: 10+	Lost	0.800	2.253	0.411
Shadab Khan, Over: 3, Economy: 10+	Wickets: 0-1, Lost	0.800	2.813	0.394
Shadab Khan, Over: 4, Sri Lanka	1st Inning, Lost	0.800	5.765	0.393
Shadab Khan, 2nd Inning, Over: 4, Wickets: 0-1, Economy: 0-6	Lost	0.750	2.112	0.373
Shadab Khan, England	Lost	0.667	1.877	0.290
Shadab Khan, Wickets: 0-1, Economy: 10+	Lost	0.667	1.877	0.290
Shadab Khan, Wickets: 0-1, England	Lost	0.667	1.877	0.278
Shadab Khan, Sri Lanka, 1st Inning	Lost	0.667	1.877	0.278
Wickets: 0-1, Shadab Khan, 2nd Inning, Economy: 0-6	Lost	0.636	1.792	0.266
Shadab Khan, Sri Lanka, 1st Inning	Over: 4, Lost	0.667	3.551	0.248
Wickets: 0-1, Shadab Khan, 2nd Inning, New Zealand	Lost	0.625	1.760	0.242
Shadab Khan, Economy: 10+	Lost	0.600	1.690	0.223
Shadab Khan, Economy: 10+	Wickets: 0-1, Lost	0.600	2.110	0.209
Shadab Khan, Sri Lanka	Over: 4, 1st Inning, Lost	0.571	6.176	0.183
Shadab Khan, 2nd Inning, England	Lost	0.571	1.609	0.183
Shadab Khan, Sri Lanka	Lost	0.571	1.609	0.183
Wickets: 0-1, Shadab Khan, 2nd Inning, Economy: 10+	Lost	0.571	1.609	0.183
Shadab Khan, Over: 4, England	Lost	0.571	1.609	0.183
Shadab Khan, Wickets: 0-1, 2nd Inning	Lost	0.500	1.408	0.175

A. Interpretation of Rules

1) Interpretation of Rules for Babar Azam from Table I and II: 1. **Rule:** If Babar Azam scores at least 50 runs, plays against New Zealand, has a strike rate of at least 120, and bats in the top order, the team is highly likely to win. - **Interpretation:** When Babar Azam meets these criteria, the confidence in winning is 100%, which means every time these conditions are satisfied, the team wins. The lift of 1.738 indicates that this association is 1.738 times more likely than a random association.

2. **Rule:** Similar to the first rule, but playing against West Indies. - **Interpretation:** The interpretation is identical to the first rule. When Babar Azam performs well under these conditions against West Indies, the confidence in winning is 100%, with a lift of 1.738.

3. **Rule:** If Babar Azam scores at least 50 runs, plays in the 2nd inning, is in the top order, the team is likely to win. - **Interpretation:** The confidence is 92.9%, indicating a strong association. The lift of 1.614 suggests that the probability of winning in these circumstances is 1.614 times higher than the expected probability.

4. **Rule:** If Babar Azam plays against West Indies, the team is likely to win. - **Interpretation:** The confidence is 92.9%, indicating a strong association. The lift of 1.614 implies that the likelihood of winning in these conditions is 1.614 times higher than the expected probability.

5. **Rule:** If Babar Azam scores at least 50 runs in the 2nd inning, the team is likely to win. - **Interpretation:** The confidence is 92.9%, indicating a strong association. The lift of 1.614 suggests that the probability of winning in these circumstances is 1.614 times higher than the expected probability.

6. **Rule:** If Babar Azam scores at least 50 runs in the 2nd inning and has a strike rate of at least 120, the team is likely to win. - **Interpretation:** The confidence is 92.3%, indicating a strong association. The lift of 1.605 implies that the probability of winning in these conditions is 1.605 times higher than the expected probability.

7. **Rule:** If Babar Azam plays against Bangladesh and scores less than 25 runs, the team is highly likely to win, and the strike rate is less than 90. - **Interpretation:** The confidence is 100%, indicating a perfect association. The lift of 6.551 is very high, suggesting a significant increase in the likelihood of winning under these conditions.

8. **Rule:** If Babar Azam plays against West Indies, the team is likely to win, especially if he is in the top order. - **Interpretation:** The confidence is 92.9%, indicating a strong association. The lift of 3.620 suggests that the probability of winning in these conditions is 3.620 times higher than the expected probability.

9. **Rule:** Similar to the eighth rule, but scoring at least 50 runs in the 2nd inning. - **Interpretation:** The interpretation is identical to the eighth rule. When Babar Azam satisfies these conditions, the confidence in winning is 92.9%, with a lift of 3.620.

10. **Rule:** If Babar Azam scores at least 50 runs in the 2nd inning and has a strike rate of at least 120, the team is likely to win, especially if he is in the top order. - **Interpretation:** The confidence is 92.3%, indicating a strong association. The lift of 3.599 suggests that the probability of winning in these conditions is 3.599 times higher than the expected probability.

11. **Rule:** If Babar Azam has a strike rate between 90 and 120, plays in the 1st inning, and scores between 25 and 50 runs, the team is likely to lose. - **Interpretation:** The confidence is 66.7%, indicating a moderate association. The lift of 1.570 suggests that the probability of losing in these conditions is 1.570 times higher than the expected probability.

12. **Rule:** If Babar Azam plays against England, the team is likely to lose, especially if he is in the top order. - **Interpretation:** The confidence is 60%, indicating a moderate association. The lift of 4.187 suggests that the probability of losing in these conditions is 4.187 times higher than the expected probability.

13. **Rule:** If Babar Azam is in the top order and plays against England, the team is likely to lose. - **Interpretation:** The confidence is 60%, indicating a moderate association. The lift of 1.413 implies that the probability of losing in these conditions is 1.413 times higher than the expected probability.

14. **Rule:** If Babar Azam plays against England, the team is likely to lose. - **Interpretation:** The confidence is 60%, indicating a moderate association. The lift of 1.413 suggests that the probability of losing in these conditions is 1.413 times higher than the expected probability.

15. **Rule:** If Babar Azam plays against Sri Lanka, the team is likely to lose. - **Interpretation:** The confidence is 62.5%, indicating a moderate association. The lift of 1.472 suggests that the probability of losing in these conditions is 1.472 times higher than the expected probability.

16. **Rule:** If Babar Azam plays against Sri Lanka, is in the top order, the team is likely to lose. - **Interpretation:** The confidence is 62.5%, indicating a moderate association. The lift of 1.472 implies that the probability of losing in these conditions is 1.472 times higher than the expected probability.

17. **Rule:** If Babar Azam plays against Sri Lanka, the team is likely to lose, especially if he is in the top order. - **Interpretation:** The confidence is 62.5%, indicating a moderate association. The lift of 4.361 suggests that the probability of losing in these conditions is 4.361 times higher than the expected probability.

18. **Rule:** If Babar Azam has a strike rate of at least 120 and plays against England, the team is likely to lose, especially if he is in the top order. - **Interpretation:** The confidence is 60%, indicating a moderate association. The lift of 4.187 suggests that the probability of losing in these conditions is 4.187 times higher than the expected probability.

19. **Rule:** If Babar Azam plays in the 1st inning against England, the team is likely to lose, especially if he is in the top order. - **Interpretation:** The confidence is 60%, indicating a moderate association. The lift of 4.187 suggests that the probability of losing in these conditions is 4.187 times higher than the expected probability.

20. **Rule:** If Babar Azam plays against England with a strike rate of at least 120, the team is likely to lose. - **Interpretation:** The confidence is 60%, indicating a moderate association. The lift of 1.413 suggests that the probability of losing in these conditions is 1.413 times higher than the expected probability.

2) Interpretation of Rules for Shadab Khan from Table III and IV: 1. **Rule:** If Shadab Khan bowls more than 4 overs against West Indies, the team is highly likely to win. - **Interpretation:** The confidence is 100%, indicating a perfect association. The lift of 1.551 suggests that the probability of

winning in these conditions is 1.551 times higher than the expected probability.

2. **Rule:** If Shadab Khan plays against West Indies, has an economy of 0-6, the team is highly likely to win. - **Interpretation:** The confidence is 100%, indicating a perfect association. The lift of 1.551 suggests that the probability of winning in these conditions is 1.551 times higher than the expected probability.

3. **Rule:** If Shadab Khan plays against Bangladesh, the team is highly likely to win. - **Interpretation:** The confidence is 100%, indicating a perfect association. The lift of 1.551 suggests that the probability of winning in these conditions is 1.551 times higher than the expected probability.

4. **Rule:** If Shadab Khan bowls more than 4 overs against West Indies with an economy of 0-6, the team is highly likely to win. - **Interpretation:** The confidence is 100%, indicating a perfect association. The lift of 1.551 suggests that the probability of winning in these conditions is 1.551 times higher than the expected probability.

5. **Rule:** If Shadab Khan takes 2-3 wickets with an economy of 6-8, the team is highly likely to win. - **Interpretation:** The confidence is 100%, indicating a perfect association. The lift of 1.551 suggests that the probability of winning in these conditions is 1.551 times higher than the expected probability.

6. **Rule:** If Shadab Khan bowls in the 1st inning with an economy of 0-6, the team is likely to win. - **Interpretation:** The confidence is 91.7%, indicating a strong association. The lift of 1.421 suggests that the probability of winning in these conditions is 1.421 times higher than the expected probability.

7. **Rule:** If Shadab Khan plays in the 2nd inning against West Indies and bowls more than 4 overs, the team is likely to win, especially if the economy is 0-6. - **Interpretation:** The confidence is 100%, indicating a perfect association. The lift of 3.973 suggests that the probability of winning in these conditions is 3.973 times higher than the expected probability.

8. **Rule:** If Shadab Khan plays against Bangladesh with an economy of 6-8, the team is likely to win, especially if he bowls more than 4 overs. - **Interpretation:** The confidence is 100%, indicating a perfect association. The lift of 2.509 suggests that the probability of winning in these conditions is 2.509 times higher than the expected probability.

9. **Rule:** If Shadab Khan takes 2-3 wickets, bowls in the 1st inning with an economy of 6-8, the team is likely to win, especially if he bowls more than 4 overs. - **Interpretation:** The confidence is 100%, indicating a perfect association. The lift of 2.509 suggests that the probability of winning in these conditions is 2.509 times higher than the expected probability.

10. **Rule:** If Shadab Khan takes 2-3 wickets against West Indies with an economy of 0-6, the team is likely to win, especially if he bowls more than 4 overs. - **Interpretation:** The confidence is 100%, indicating a perfect association. The lift of 2.509 suggests that the probability of winning in these conditions is 2.509 times higher than the expected probability.

11. **Rule:** If Shadab Khan bowls more than 4 overs in the 1st inning against Sri Lanka, the team is likely to lose. - **Interpretation:** The confidence is 80%, indicating a strong association. The lift of 2.253 suggests that the probability of losing in these conditions is 2.253 times higher than the expected probability.

12. **Rule:** If Shadab Khan bowls more than 4 overs against Sri Lanka, the team is likely to lose. - **Interpretation:** The confidence is 80%, indicating a strong association. The lift of 2.253 suggests that the probability of losing in these conditions is 2.253 times higher than the expected probability.

13. **Rule:** If Shadab Khan bowls 3 overs with an economy of 10+, the team is likely to lose. - **Interpretation:** The confidence is 80%, indicating a strong association. The lift of 2.253 suggests that the probability of losing in these conditions is 2.253 times higher than the expected probability.

14. **Rule:** If Shadab Khan takes 0-1 wickets, bowls 3 overs with an economy of 10+, the team is likely to lose. - **Interpretation:** The confidence is 80%, indicating a strong association. The lift of 2.253 suggests that the probability of losing in these conditions is 2.253 times higher than the expected probability.

15. **Rule:** If Shadab Khan bowls 3 overs with an economy of 10+ and takes 0-1 wickets, the team is likely to lose. - **Interpretation:** The confidence is 80%, indicating a strong association. The lift of 2.813 suggests that the probability of losing in these conditions is 2.813 times higher than the expected probability.

16. **Rule:** If Shadab Khan bowls more than 4 overs against Sri Lanka, especially in the 1st inning, the team is likely to lose. - **Interpretation:** The confidence is 80%, indicating a strong association. The lift of 5.765 suggests that the probability of losing in these conditions is 5.765 times higher than the expected probability.

17. **Rule:** If Shadab Khan plays in the 2nd inning, bowls more than 4 overs, takes 0-1 wickets, and has an economy of 0-6, the team is likely to lose. - **Interpretation:** The confidence is 75%, indicating a strong association. The lift of 2.112 suggests that the probability of losing in these conditions is 2.112 times higher than the expected probability.

18. **Rule:** If Shadab Khan plays against England, the team is likely to lose. - **Interpretation:** The confidence is 66.7%, indicating a moderate association. The lift of 1.877 suggests that the probability of losing in these conditions is 1.877 times higher than the expected probability.

19. **Rule:** If Shadab Khan takes 0-1 wickets, has an economy of 10+, the team is likely to lose. - **Interpretation:** The confidence is 66.7%, indicating a moderate association. The lift of 1.877 suggests that the probability of losing in these conditions is 1.877 times higher than the expected probability.

20. **Rule:** If Shadab Khan takes 0-1 wickets against England, the team is likely to lose. - **Interpretation:** The confidence is 66.7%, indicating a moderate association. The lift of 1.877 suggests that the probability of losing in these conditions is 1.877 times higher than the expected probability.

V. Conclusion: Decoding Cricket Dynamics Through Player Performance Analysis

The exploration of association rules generated by the Apriori algorithm, coupled with the calculation of a holistic significance score, has provided a nuanced understanding of the intricate dynamics of cricket. This dual-player analysis delved into the performances of Babar Azam and Shadab Khan, unraveling complex patterns that significantly influence a cricket team's outcomes. The extensive examination of these association rules not only illuminates specific performance scenarios but also introduces a novel scoring mechanism that encapsulates the multifaceted nature of player contributions.

A. Association Rules and Significance Score

Table I and II, focusing on batsman Babar Azam, unravel a tapestry of conditions under which team victories are highly probable. The significance of each rule is quantified by a score, a composite of confidence, lift, Jaccard's coefficient, and Kulczynski measure. The weighting of each metric in the scoring process was meticulously determined through iterative testing, reflecting the nuanced importance of each evaluation criterion. Notably, the utilization of min-max normalization ensured equitable treatment of attributes, mitigating the impact of the broad lift value range.

B. Findings and Contributions

The association rules for Babar Azam underscore the criticality of metrics such as runs scored, strike rate, and batting order in shaping favorable outcomes. The interplay of these factors reveals strategic nuances that can guide team management in optimizing player roles and team compositions for success.

C. Shadab Khan's Bowling Dynamics

Tables III and IV pivot the analysis toward bowler Shadab Khan, shedding light on conditions favoring team victories in bowling scenarios. The scoring mechanism unveils the significance of Shadab Khan's prowess in constraining opponents, particularly against specific teams and in specific innings.

D. Implications and Significance

The implications of these findings extend beyond individual player statistics, offering actionable insights for team management, coaches, and strategists. The contextual conditions highlighted in the association rules provide a strategic roadmap for optimizing team compositions and gameplay strategies based on the strengths of key players.

E. Limitations and Future Research

It is imperative to acknowledge the limitations of this study, including its retrospective nature and reliance on historical performance data. Future research endeavors could explore real-time analytics, incorporating dynamic variables such as weather conditions, player form fluctuations, and evolving team strategies.

F. Concluding Remarks

In conclusion, the synergistic analysis of Babar Azam and Shadab Khan's performances showcases the profound impact individual players can have on team outcomes. The association rules, along with the innovative scoring mechanism, contribute to the evolving field of cricket analytics, providing a nuanced understanding of player dynamics and their influence on match outcomes. As cricket analytics continues to evolve, the findings from this analysis serve as a foundation for future research, emphasizing the pivotal role of data-driven insights in shaping strategic decisions and fostering success in the dynamic realm of international cricket. This work marks a significant stride in unraveling the complexity of cricket dynamics and adds a valuable dimension to the growing field of sports analytics.

References

[1] Sanmoy Bhattacherjee, Jayakrushna Sahoo, Adrijit Goswami "Association Rule Mining Approach in Strategy Planning for Team India in ICC World Cup 2015," Department of Mathematics, IIT Kharagpur,Kharagpur, India.

[2] Utsav Jagdishbhai Solanki, Prof. Jay Vala, "Selection for Balanced Cricket Team Fourth coming ICC Championship 2017", G.H.Patel Collage of Engineering and Technology, V.V.Nagar, Anand, Gujarat, India.

[3] Dataset Downloaded from Kaggle link https://www.kaggle.com/datasets/jamiewelsh2/ball-by-ball-it20/data.

[4] Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD international conference on Management of data.

[5] Witten, I. H., Frank, E., and Hall, M. A. (2016). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.

[6] M. Lewis, *Moneyball: The Art of Winning an Unfair Game*. W. W. Norton & Company, 2004.

[7] B. C. Alamar, *Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers*. Columbia University Press, 2013.

[8] I. M. Franks, *Football Analytics: Science, Performance Analysis and Competitive Advantage*. Routledge, 2016.

APPENDIX

Cricket is a globally popular sport played between two teams, with each team traditionally comprising 11 players. The game involves a unique interplay between bat and ball and is recognized for its rich history and cultural significance, particularly in countries such as England, Australia, India, and numerous others.

A. Basic Rules

1) **Teams:** Each team consists of batsmen, bowlers, and fielders.
2) **Innings:** A match is divided into innings, with each team batting and bowling.
3) **Runs:** Batsmen score runs by hitting the ball and running between wickets.
4) **Dismissals:** Batsmen can be dismissed in various ways, such as getting bowled, caught, or leg-before-wicket (LBW).
5) **Bowling:** Bowlers aim to dismiss batsmen and limit their runs.

B. Team Composition

In a standard cricket match, each team fields 11 players, and the composition of the team is critical, encompassing various player roles such as batsmen, bowlers, all-rounders, and wicketkeepers.

C. Objective

The primary objective in cricket is for a team to outscore its opponent. This is achieved by accumulating runs through a combination of batting and running between two sets of stumps at either end of the pitch. A run is scored when the batsmen successfully traverse the length of the pitch, exchanging positions.

D. Scoring

Cricket offers various ways for a team to score runs, including hitting the ball beyond the boundary (resulting in a boundary or a six), running between the wickets, and penalties incurred by the opposing team. Additionally, bowlers aim to dismiss batsmen through various means, such as getting them out bowled, caught, or leg-before-wicket (LBW).

E. Innings

A cricket match is typically divided into innings, with each team having the opportunity to bat and bowl. The team batting aims to set a high score, while the bowling team endeavors to dismiss the opposing batsmen and restrict the runs scored.

Cricket is a dynamic and strategic sport played on an oval-shaped field, with the outcome determined by a combination of individual and team performances in batting, bowling, and fielding. The game's intricate rules and nuances contribute to its global appeal and cultural significance.

F. Formats

- **Test Cricket:** Traditional, played over five days.Due to its extended play, test matches are known for their endurance and strategic gameplay. Teams have the opportunity to recover from setbacks and showcase their resilience over the extended duration.Test cricket is often regarded as the ultimate test of a player's skills, as it demands not only technical prowess but also mental stamina. The format allows for more nuanced strategies, and weather conditions can significantly impact the course of the match.
- **One Day Internationals (ODIs):** Limited to 50 overs per side.ODIs are completed in a single day, making them more accessible for both players and spectators. The limited overs create an environment where teams need to balance scoring quickly with preserving wickets.ODIs are known for their dynamic and often high-scoring nature. Batsmen aim to score at a brisk pace, and bowlers strive to take wickets to restrict the opposing team's total. This format is popular in international and domestic cricket.
- **Twenty20 (T20):** A fast-paced format with each team facing 20 overs.T20 matches are designed to be completed in approximately three to four hours, making them highly suitable for shorter attention spans and maximizing excitement.T20 cricket emphasizes aggressive and innovative play. Batsmen often take more risks, and bowlers focus on delivering variations to outsmart the opposition. The format has gained immense popularity, especially in domestic leagues and international tournaments.

G. Association Rule Mining

Association rule mining is a data mining technique used to discover interesting relationships, patterns, or associations within large datasets. In the context of cricket analytics, association rules can reveal dependencies between different factors and outcomes.

H. Apriori Algorithm

The Apriori algorithm is commonly used for association rule mining. It identifies frequent itemsets and generates rules based on their occurrences in the dataset.

I. Confidence and Lift

- **Confidence:** Measures the likelihood of the consequent (output) occurring given the antecedent (input).
- **Lift:** Represents the ratio of the observed support to that expected if the antecedent and consequent were independent.

The score for each association rule is calculated using a weighted combination of four evaluation metrics: confidence, lift, Jaccard's coefficient, and Kulczynski measure. The steps involved are as follows:

1) **Normalization:** Normalize each metric's values to bring them to a comparable scale.
2) **Weights:** Assign weights to the normalized attributes based on their significance.

3) **Weighted Score:** Multiply each normalized attribute by its weight and sum the results to obtain the overall score.

This scoring mechanism ensures that each metric contributes proportionally to the final score, reflecting the importance of different aspects in evaluating the association rules.

By applying these methodologies, the association rules provide valuable insights into the conditions under which a cricket team is likely to win or lose, contributing to a deeper understanding of the game's dynamics.

In addition to the detailed explanation of cricket rules and association rule mining, it's important to acknowledge some limitations of the analysis:

- The quality of predictions depends on the quality and completeness of the data.
- The assumptions made during rule generation may not capture all nuances of the game.
- External factors such as weather conditions or player injuries are not considered.

J. Suggestions for Future Research

To enhance the predictive power and robustness of cricket analytics, future research could explore:

- Integration of real-time data for more dynamic insights.
- Inclusion of player-specific attributes for a personalized analysis.
- Evaluation of association rules in the context of different cricket formats.

These considerations and suggestions aim to inspire further exploration and refinement of cricket analytics methodologies.