Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 7 (2025)

Probabilistic Fuzzy Classification for Strategic Player Selection in T20I Cricket

[1]Muhammad Waqas, [2*]Qamruz Zaman, [3]Imran ullah, [4]Haleema Anwar

Article Details

Keywords: Fuzzy C Means (FCM), T-20 Cricket, Probabilistic, Player Selection, Membership Function

Muhammad Waqas
Department of Statistics, University of Peshawar, Pakistan.
waqasuop1416@gmail.com
Qamruz Zaman
Department of Statistics, University of Peshawar, Pakistan. Corresponding Author Email: cricsportsresearchgroup@gmail.com
Imran ullah
Institute of Numerical Sciences (INS) Kohat University of Science and Technology.
imran0310103@gmail.com
Haleema Anwar
Institute of Numerical Sciences (INS) Kohat University of Science and Technology.
Haleemaanwar63@gmail.com

ABSTRACT

This study presents a knowledge-based intelligent system for T20I cricket team selection using probabilistic fuzzy clustering. Utilizing performance data from 336 international players spanning 2005 to 2024, the analysis classifies players into four functional categories: specialist batters, specialist bowlers, balanced all-rounders, and bowling all-rounders. The Fuzzy C-Means (FCM) clustering algorithm, integrated with Principal Component Analysis (PCA), is employed to handle the multidimensional nature of player statistics while capturing the uncertainty and role overlap typical in T20 cricket. The findings indicate that fuzzy clustering effectively identifies ideal substitutes and balanced team compositions by assigning probabilistic memberships across multiple roles. Eight representative players are selected from each cluster, illustrating the model's capacity to support strategic, flexible, and data-driven team selection. This probabilistic fuzzy logic framework provides a transparent and practical solution for optimizing squad composition in the dynamic context of T20 cricket..

INTRODUCTION

In sports, success is determined not only by decisions made during the game but also by those made beforehand. Factors such as team selection, training regimes, and team-building efforts are just as crucial to performance as tactical decisions and on-field execution [1]. Squad selection is one of the most debated topics in cricket, frequently discussed in the media and among fans. The process requires a blend of expertise, experience, and intuition. A significant portion of the cricketing community often expresses the need for greater transparency in the decisions made by national selection committees, which are tasked with choosing players for international representation. One of the major challenges in cricket analytics is identifying the most suitable replacements for players. An effective replacement must deliver comparable performance levels while maintaining team cohesion. Ideally, the new player should mirror the outgoing player in terms of skill, playing style, and historical performance metrics.[2,3]. Fuzzy modeling is an essential topic in fuzzy logic, generally classified into two main categories. The first involves fuzzy relational equations, which closely align with the structure of traditional state-space formulations [1]. Among the various techniques in fuzzy modeling, fuzzy clustering is widely adopted. It offers the benefit of automatically generating a fuzzy partition of the input or product space. Researchers frequently use clustering methods to locate patterns in the product space and then construct fuzzy partitions of the input space by mapping these clusters [2]. It has played a significant role in pattern recognition, data mining, and image processing [3]. In recent decades, many clustering techniques with strong performance have been introduced, which are generally grouped into two categories: hard clustering and fuzzy (or soft) clustering [4]. Hard clustering assigns each data point to exactly one cluster, whereas fuzzy clustering allows data points to belong to multiple clusters, each with a certain membership value [5]. In today's digital era, the rapid and exponential growth of data in both size and dimension has outpaced the capabilities of traditional clustering methods. This makes it critically important to develop more efficient and scalable clustering algorithms to manage large-scale data challenges effectively [6].

In applied settings, it is common for a single data instance to belong to multiple clusters simultaneously, making fuzzy logic particularly well-suited for representing such inherent ambiguity. The theoretical foundation for this approach was established by Zadeh through the introduction of fuzzy set theory [7]. Building on this framework, Bezdek introduced the fuzzy c-means (FCM) algorithm, a groundbreaking advancement in fuzzy clustering methodologies. FCM is not only computationally efficient but also adept at capturing uncertainty within data.

Numerous refined versions of FCM have since been developed to enhance performance and adaptability. Unlike hard clustering methods that enforce strict classification, fuzzy clustering through FCM and its derivatives produces soft partitions, offering a more flexible and informative interpretation of complex data structures [5]. The assignment of membership values in fuzzy clustering is highly susceptible to noise, often leading to degraded clustering accuracy. To counter this issue, significant research has been devoted to developing techniques aimed at reducing the impact of noise on clustering outcomes [8, 9, and 10]. To resolve the issue of a data point being equidistant from multiple prototypes, Krishnapuram and Keller [23] proposed the Possibilistic C-Means (PCM) algorithm. PCM departs from traditional fuzzy clustering by relaxing the strict membership constraints, embodying the essence of the probabilistic framework. Building upon this concept, the probabilistic Fuzzy C-Means (PFCM) algorithm [9] was later introduced, allowing for the concurrent estimation of both membership degrees and possibility values.

This study proposed a knowledge-based intelligent system that utilizes probabilistic fuzzy clustering to recommend substitutes for team selection in T20I cricket. It serves as a practical and efficient tool for identifying ideal player replacements.

METHODOLOGY OF THE STUDY

The methodology of this study is grounded in a probabilistic fuzzy logic approach aimed at optimally classifying and selecting T20 International (T20I) cricket players based on their batting and bowling performance metrics.

The fuzzy clustering procedure carried out using the **fuzzy c-means clustering** algorithm minimizes the following **objective function**

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \cdot \|x_i - c_j\|^2$$

Where;

- N: Total number of data points (players)
- C: Number of clusters
- x_i: i^{th} data point (feature vector of a player)
- c_j: Center of the j^{th} cluster
- u_{ij}: Membership degree of x_i in cluster j
- m: Fuzziness coefficient (typically $m = 2$)
- $||x_i - c_j||^2$: Squared Euclidean distance between data point x_i and cluster center c_j

Also,

Cluster Centers Update:

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m}$$

Membership Update:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{||x_i - c_j||}{||x_i - c_k||} \right)^{\frac{2}{m-1}}}$$

DATA COLLECTION

A comprehensive dataset was compiled from the official website of Cricbuzz.com encompassing detailed performance statistics of all bowlers, batsmen, and all-rounders across international matches played between 2005 and 2024. In present study, the total of 336 players were initially considered by combining two datasets, one containing batting statistics and the other bowling statistics. These datasets included critical performance indicators such as batting average, strike rate, total runs, centuries, half-centuries, number of fours and sixes, bowling average, economy rate, strike rate (balls per wicket), and average wickets per match.

After merging the datasets using unique player identifiers, the data was cleaned and filtered to retain only those players with complete numeric records, resulting in a final sample of 199 players. All selected variables were standardized using z-score normalization to ensure comparability across different statistical ranges. The study employed fuzzy c-means clustering, a soft clustering method that assigns players to multiple clusters with varying degrees of membership rather than categorizing them rigidly. This approach was ideal for capturing the overlapping nature of player roles in T20 cricket, where individuals often contribute across both batting and bowling domains. The flow chart of the procedure adopted in fuzzy clustering for team selection is depicted in Figure 1.

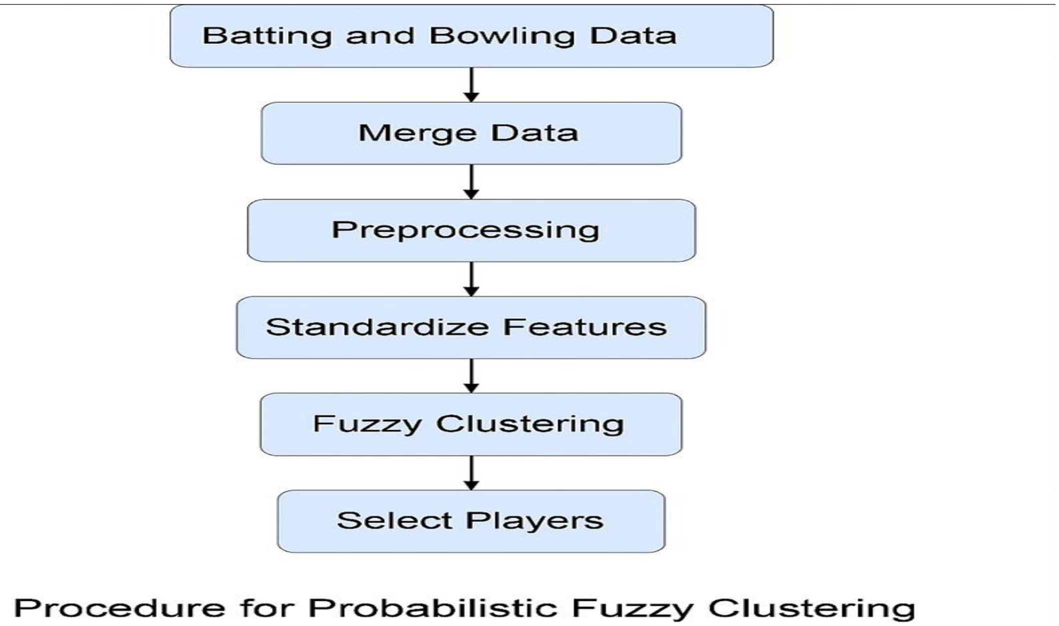

FIGURE 1: PROCEDURE FOR PROBABILISTIC FUZZY CLUSTERING FOR TEAM SELECTION IN T20I CRICKET

Based on domain knowledge and performance diversity, the number of clusters was pre-set to four, representing likely roles such as specialist batters, specialist bowlers, balanced all-rounders, and bowling all-rounders or utility players. To aid in interpretation and visualization, Principal Component Analysis (PCA) was applied to reduce the multi-dimensional performance data into two principal components, enabling graphical representation of clusters and highlighting the probabilistic boundaries between them. The methodology provides a structured framework for understanding player roles and forming balanced team compositions based on objective performance criteria and probabilistic role affiliations.

RESULTS AND DISCUSSION

FUZZY CLUSTERING FOR TEAM SELECTION

The figure 2 represents the fuzzy clustering results of T20 cricket players based on a comprehensive set of batting and bowling performance metrics, projected onto a two-dimensional space using Principal Component Analysis (PCA) for visualization. Each point in the scatter plot corresponds to an individual player, and the shape and color of the point indicate the player's dominant cluster membership as determined by the fuzzy c-means algorithm. Unlike hard clustering methods, fuzzy clustering allows players to partially belong to multiple clusters, making it particularly well-suited for evaluating multifaceted player roles in T20 cricket.

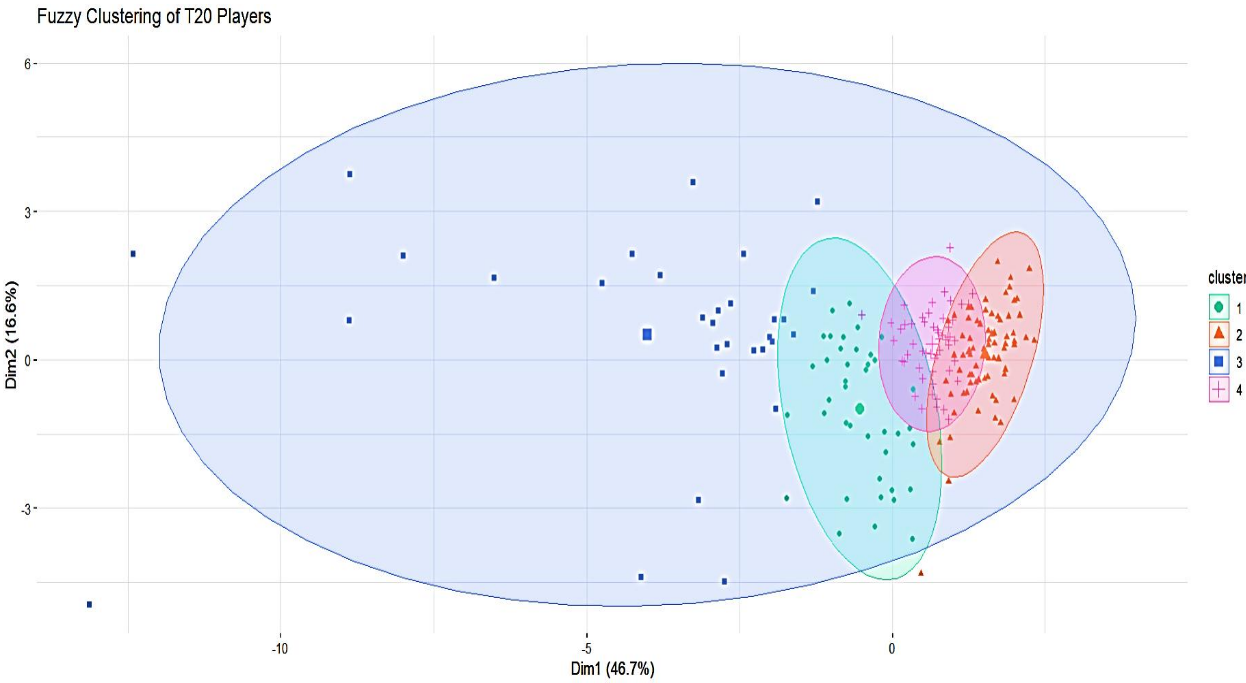

FIGURE 2: VISUAL DIAGRAM OF FUZZY CLUSTERING FOR TEAM SELECTION IN T20I CRICKET

In this visualization:

- Cluster 1 (green circles) appears primarily in the central-left portion of the graph and represents players who are balanced or true all-rounders, contributing significantly with both bat and ball. These players display moderate to high batting strike rates and averages, along with reliable bowling metrics, making them versatile assets in T20 teams.
- Cluster 2 (orange triangles) forms a tighter, more compact group, suggesting players who are specialists with highly similar performance profiles. Depending on the underlying stats, this cluster may include top-order batters or frontline bowlers who exhibit focused excellence in a single discipline.
- Cluster 3 (blue squares) is more widely scattered and generally located further to the left of the plot. This dispersion suggests greater variability or inconsistency within the group, which may include low-performing players, outliers, or emerging players with incomplete or unbalanced contributions.
- Cluster 4 (pink crosses) overlaps portions of Clusters 1 and 2 and includes players whose performance characteristics span multiple roles. These are likely utility players or bowling all-rounders who exhibit strength in one domain and moderate competence in the other, providing depth and adaptability to team combinations.

The ellipses surrounding each cluster denote the soft probabilistic boundaries of cluster membership. A player's proximity to the center of a cluster implies a higher degree of membership, while players located on the borders may share characteristics with multiple clusters. This soft assignment allows for nuanced understanding of roles in a format like T20, where versatility is highly valued.

The two principal components, Dim1 and Dim2, capture 46.7% and 16.6% of the total variance, respectively. While not exhaustive, this cumulative 63.3% variance is sufficient to meaningfully differentiate player types and highlight key performance-based patterns. The PCA transformation reduces the multidimensional statistical space into a more interpretable 2D form, aiding analysts, coaches, and selectors in visually assessing how players are grouped and what roles they are statistically inclined to perform.

Overall, this figure serves as a powerful tool for team management, enabling data-driven selection strategies by identifying clusters of similar players and assessing overlap or uniqueness in player contributions. It reflects the dynamic and probabilistic nature of modern cricket analytics, particularly suited to the fast-paced, all-format demands of T20 cricket.

DESCRIPTION OF FUZZY CLUSTERING FOR TEAM SELECTION IN T20I CRICKET

Based on the numeric summary of the fuzzy clusters in Table 1, the players in each group can be interpreted according to their batting and bowling performances. Cluster 1 represents *all-rounders* who contribute decently with both bat and ball, reflected by a moderate batting average (21.9), a good strike rate (132), and solid bowling stats including 17.1 wickets and an economy of 5.62. Cluster 2 comprises *specialist bowlers* with minimal batting ability (average 7.12, SR 79.5) but strong bowling performance—taking 50.8 wickets with an excellent economy (5.52) and the best strike rate (35.2).

TABLE 1: CLUSTER-WISE PERFORMANCE SUMMARY OF T20 PLAYERS

Cluster	Batting Avg	Strike Rate (SR)	Runs	Wickets	Economy	Bowling Strike Rate
1	21.90	132.00	476	17.1	5.62	55.7
2	7.12	79.50	47.5	50.8	5.52	35.2
3	28.50	133.00	1752	42.2	5.52	53.3
4	17.20	132.00	198	54.2	5.30	36.9

Cluster 3 is characterized by *top batters*, exhibiting the highest batting average (28.5), strike rate (133), and total runs (1752), while still offering moderate bowling contributions. Cluster 4 appears to contain *bowling all-rounders or utility players* those with decent lower-order batting (average 17.2, SR 132) and very strong bowling (54.2 wickets, economy 5.30, and strike rate 36.9). Overall, the fuzzy clustering has effectively grouped players into meaningful roles that balance batting and bowling strengths, offering a strategic foundation for T20 team selection.

SELECTION OF PLAYERS BASED ON FUZZY CLUSTERING

The fuzzy clustering model applied to batting and bowling statistics has identified eight players from four distinct clusters in Table 2, each representing a specific player profile in T20 cricket. From Cluster 1, players with IDs 465793 and 345821 have been selected. This cluster consists of balanced all-rounders, players who show a reasonable batting average and strike rate, coupled with effective bowling contributions. These players are valuable assets in T20 teams due to their ability to contribute in both departments, offering flexibility in team composition.

From Cluster 2, players 552152 and 538506 have emerged as top members. This cluster is characterized by specialist bowlers players who may have minimal impact with the bat but possess outstanding bowling metrics such as high wicket counts, excellent economy rates, and sharp strike rates. Their role is crucial in controlling runs and taking key wickets during critical phases of the game.

TABLE 2: SELECTED PLAYER BY FUZZY CLUSTERING

Player ID	Player Name	Cluster	Cluster Role
465793	DM de Silva	1	Balanced All-Rounder
345821	MD Gunathilaka	1	Balanced All-Rounder
552152	PVD Chameera	2	Specialist Bowler
538506	Taskin Ahmed	2	Specialist Bowler
56025	Mahmudullah	3	Specialist Batter
55870	SC Williams	3	Specialist Batter
541224	LM Jongwe	4	Bowling All-Rounder
784379	PWH de Silva	4	Bowling All-Rounder / Utility

In Cluster 3, players 56025 and 55870 are selected, representing the group of elite batters. These players demonstrate the highest averages and run aggregates among all clusters, with aggressive strike rates. Their selection reflects their capability to anchor the innings or accelerate scoring when needed. Though their bowling contributions are limited or secondary, their primary role is to dominate with the bat.

Finally, Cluster 4 includes players 541224 and 784379, identified as bowling all-rounders or utility players. These individuals have excellent bowling records—leading in wickets and maintaining strong economy rates, while also providing moderate support with the bat. They typically bat lower in the order but are capable of contributing important runs and turning matches with the ball.

Together, these eight players cover the essential roles needed for a well-rounded T20 side: specialist batters, frontline bowlers, and dual-role all-rounders. The fuzzy clustering approach ensures that selection is based not only on rigid roles but also on probabilistic membership, acknowledging that players may contribute across multiple areas. This makes the selection process more nuanced and adaptive to dynamic team strategies.

CONCLUSION

This research highlights the effectiveness of probabilistic fuzzy clustering in addressing the complexities of team selection in T20I cricket. By leveraging detailed performance data and applying fuzzy C-means clustering, the study successfully identified and categorized players into four key roles: specialist batters, specialist bowlers, balanced all-rounders, and bowling all-rounders. The integration of Principal Component Analysis allowed for better visualization and interpretation of player groupings. Unlike traditional selection methods, the fuzzy logic approach accommodates the overlapping nature of player roles, offering a more flexible and realistic evaluation. The final selection of eight players from diverse clusters demonstrates the model's practical utility in forming balanced and strategically sound teams. Overall, this intelligent, data-driven framework offers a robust, transparent, and adaptable solution for cricket selectors, enhancing the overall quality and consistency of T20I team compositions.

REFERENCES

[1] Das, N. R., Ghosh, S., Mukherjee, I., & Paul, G. (2023). Adoption of a ranking based indexing method for the cricket teams. *Expert Systems with Applications*, *213*, 118796.

[2] Kimber, A. C., & Hansford, A. R. (1993). A statistical analysis of batting in cricket. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *156*(3), 443-455.

[3] Damodaran, U. (2006). Stochastic dominance and analysis of ODI batting performance: The Indian cricket team, 1989-2005. *Journal of sports science & medicine*, *5*(4), 503.

[4]. Pedrycz, W. (1984). An identification algorithm in fuzzy relational systems. *Fuzzy sets and systems*, *13*(2), 153-167.

[5] Klawonn, F., & Kruse, R. (1997). Constructing a fuzzy controller from data. *Fuzzy sets and systems*, *85*(2), 177-193.

[6] Nayak, J., Rekha, H. S., & Naik, B. (2023). Fuzzy C-Means Clustering: Advances and Challenges (Part II). *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, 239-269.

[7] Bhardwaj, M., Khan, N. U., & Baghel, V. (2025). A Novel Fuzzy C-Means Clustering Framework for Accurate Road Crack Detection: Incorporating Pixel Augmentation and Intensity Difference Features. *Informatica*, *49*(15).

[8] Bhardwaj, M., Khan, N. U., & Baghel, V. (2024). Fuzzy C-Means clustering based selective edge enhancement scheme for improved road crack detection. *Engineering Applications of Artificial Intelligence*, *136*, 108955.

[9] Zhou, L., Du, G., Lü, K., Wang, L., & Du, J. (2024). A survey and an empirical evaluation of multi-view clustering approaches. *ACM Computing Surveys*, *56*(7), 1-38.

[10] Dzitac, I. (2021). Zadeh's centenary. *International Journal of Computers Communications & Control*, *16*(1).

[11] Chen, J., Zhang, H., Pi, D., Kantardzic, M., Yin, Q., & Liu, X. (2021). Research Article A Weight Possibilistic Fuzzy C-Means Clustering Algorithm.

[12] Chen, J., Zhu, J., Jiang, H., Yang, H., & Nie, F. (2022). Sparsity fuzzy C-means clustering with principal component analysis embedding. *IEEE Transactions on Fuzzy Systems*, *31*(7), 2099-2111.

[13] Mei, S., Li, F., Gao, Q., & Yang, M. (2024). Anchor-free Clustering based on Anchor Graph Factorization. *arXiv preprint arXiv:2402.15688*.