Journal of Theoretical and Applied Information Technology
31st January 2025. Vol.103. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MACHINE LEARNING APPROACHES IN FANTASY CRICKET: COMPARATIVE EVALUATION OF USER-CREATED, RANDOM, AND K-MEANS CLUSTERING METHODS

POLINATI VINOD BABU [1,2], Dr. M.V.P. CHANDRA SEKHARA RAO [3]

[1]Research Scholar, Acharya Nagarjuna University, Department of CSE, India

[2]Assistant Professor, Shri Vishnu Engineering College for Women, Department of AI, India

[3]Professor, RVR & JC College of Engineering, Department of CSE, India

E-mail: [1]vinodbabusir@gmail.com, [2]manukondach@gmail.com

ABSTRACT

This research examines three approaches to building fantasy cricket teams: User-Created Procedure, Random Procedure, and K-Means Clustering Algorithm. The objective is to identify the best formation strategy by examining player performance data from six games. While the Random Procedure creates teams by picking players at random within predetermined parameters, the User-Created Procedure uses manual selection based on intuitive strategies. Using a machine learning technique, the K-Means Clustering Algorithm groups teams according to credit and performance indicators to find the best-performing teams that stay within credit restrictions. This optimizes team formation. Our findings show that, in terms of overall performance, user-created and randomly generated teams are regularly out performed by the K-Means Clustering technique. This study demonstrates how machine learning techniques can improve the development of fantasy cricket teams by providing a data-driven method that is superior to traditional and random approaches.

Keywords: *K-Means Clustering, Fantasy Cricket, Fantasy Points System, Team Optimization, and Player Performance.*

1. INTRODUCTION

Fantasy cricket has enjoyed significant popularity over the last two decades. This online game has sparked a new trend in the cricket-loving countries. Players construct fictitious teams of genuine players and receive points based on their performance in real matches. Fantasy cricket has an impact on more than just entertainment; it affects sports, the economy, technology, and interactions with others.

The popularity of fantasy cricket platforms such as Dream11, My11Circle, and MPL has increased rapidly. In 2019, Dream11, the trailblazer, attained unicorn status. Fantasy cricket leagues are played by millions of users from a wide range of age groups and demographics. Over 100 million users were active on fantasy sports platforms during the 2023 Indian Premier League (IPL). The fantasy cricket market has grown to be a profitable sector of the economy. Platforms get revenue from partnerships, admission fees, and advertisements. Numerous career possibilities in tech, marketing, data analysis, and customer service have been generated by the industry. The financial ecosystem has been improved by the sector's significant attraction of venture capitalist investments and strategic alliances with important cricket competitions.

Data analytics has advanced thanks in large part to the popularity of fantasy cricket. To make well-informed decisions, users depend on comprehensive information and performance measures. Artificial Intelligence (AI) and Machine Learning algorithms are used by platforms to improve user experience, forecast player performance, and stop fraud. With real-time updates and fluid game play, mobile application and user interface innovations have made it simpler for fans to interact with fantasy cricket. Cricket fans can develop a sense of community through fantasy cricket. Colleagues, friends, and family organize leagues, have conversations, and share in victories. The popularity of fantasy cricket has raised cricket audience engagement. Fans keep track of individual players' performances in a variety of matches and competitions in addition to their favorite teams. By helping players gain a deeper understanding of player data, match conditions, and strategies, the game helps users become more knowledgeable cricket fans.

Many times, people criticize fantasy cricket for encouraging gambling and want tighter laws and moral considerations. Because of the game's addictive qualities, players may spend too much time and money, which may have an impact on their personal and professional lives. The legal positions on fantasy sports vary across different jurisdictions, creating challenges for platforms operating in a complex regulatory environment. Unquestionably, fantasy cricket has changed the sports entertainment scene. It has a wide-ranging effect on relationships, technology, and the economy. Despite its difficulties, the sector has enormous development potential and the opportunity to positively impact cricket as a sport and its supporters. As the industry develops further, resolving moral and legal issues will be essential to maintaining its prosperity and optimizing gains for all parties.

The datasets generated and analyzed during the current study are not publicly available due to proprietary restrictions but are available from the corresponding author upon reasonable request. All data used for fantasy cricket team analysis, including player performance metrics, credits, and match outcomes, can be shared for academic research purposes, provided the necessary permissions are obtained from the data source.

2. LITERATURE SURVEY

Because of its intricacy and ability to make use of a variety of cutting-edge analytical and computational tools, fantasy cricket team selection has attracted a lot of interest. This review of the literature gives an overview of the major works in this field, emphasizing the research methods and conclusions that have advanced the field of fantasy cricket team selection.

Jha et al. (2011) [1] suggest a hybrid strategy for fantasy cricket team selection optimization that combines genetic algorithms and recursive feature elimination (RFE). The significance of taking into account different player performance metrics and match situations is emphasized by this study. The goal of the hybrid technique is to combine the advantages of genetic algorithms and RFE to improve the efficacy and accuracy of team selection. Similarly, employing recurrent neural networks and genetic algorithms, Kumarasiri & Perrera (2017) [6, 15] provide optimized models for cricket team selection. Their research focuses on using evolutionary algorithms to maximize team selection based on player performance data and budgetary constraints, while navigating the large solution space. This method seeks to efficiently strike a balance between team strength and financial constraints.

An integer programming model is used by Gerber & Sharp (2006) [2] to choose a cricket team while accounting for player performance in various game situations. The team selection problem can be structured and solved with the use of mathematical optimization approaches, as demonstrated by this work. A multi-criterion decision-making (MCDM) method is presented by Dey et al. (2011) [3] for assessing bowler performance in the Indian Premier League (IPL). This strategy helps with better squad selection judgments by taking into account a variety of performance metrics, guaranteeing a thorough evaluation of player talents. A bi-objective optimization problem for choosing cricket teams based on player performance statistics is examined by Farquhar & Meeds (2007) [9]. Their research highlights the necessity of striking a balance between a number of goals, including meeting financial obligations and optimizing team performance. A data science method for predicting the top fantasy cricket team using greedy and knapsack algorithms is described by Kumar et al. (2022) [4]. By finding the best player combinations, this study highlights how data science and analytics may improve team selection.

Machine learning techniques are presented by Iyer & Sharda (2009) [5] as a means of team prediction for fantasy leagues. Their study emphasizes how machine learning can help with team selection issues by providing data-driven insights that enhance decision-making. Juan et al. (2021) [7], [8] describe a number of machine learning methods for forecasting the performance of cricket teams and the winners of fantasy cricket

Journal of Theoretical and Applied Information Technology
31st January 2025. Vol.103. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

competitions. These studies demonstrate how machine learning models can be used to analyze historical data and predict future events, which helps users build competitive fantasy teams.

Recent scholarship examining fantasy sports, daily fantasy sports (DFS), and sports betting [10], [11], and [12] consistently highlights the interplay between social engagement, monetary incentives, and consumption behavior. For instance, a survey of 253 fantasy baseball participants revealed that those who paid league entry fees were more strongly motivated by social factors than by financial gain, a finding that contrasts with traditional gambling research linking heavier gambling with anti-social tendencies and financial motivations. Moreover, research involving over 500 DFS participants identified parallels between DFS and other gambling contexts, such as chasing behaviors and indications of problem gambling, underscoring the need for potential consumer protections. Finally, a study of 1,555 NFL fans and bettors found that while team loyalty could moderate gambling expenditures—lowering spend among highly loyal bettors—engagement in betting itself was associated with higher overall media consumption, suggesting a complementary relationship between betting activities and sport fandom.

To provide light on the social interactions and relationships that are developed through involvement, Kissane & Winslow (2016) [13] investigate gender, social interaction, and relationships in fantasy sports. A more comprehensive view of the social background of fantasy cricket is provided by this study. Fisher (2015) [14] looks into the economic effects and expenditure patterns associated with fantasy sports. Comprehending these economic aspects is essential to conducting a thorough analysis of the fantasy cricket ecology.

The literature on choosing a fantasy cricket team uses a wide range of approaches, including machine learning, social studies, hybrid algorithms, and mathematical models. All of these studies add to our understanding of the complexity of choosing a team and provide insightful information that can improve the efficiency and fun of fantasy cricket.

3. METHODOLOGY

Three efficient techniques for team analysis in fantasy cricket are K-Means Clustering, Systematic Replacements, and Random Sampling. By picking members at random according to predetermined roles and restrictions, Random Sampling creates a huge number of teams and facilitates the exploratory examination of various team combinations. By carefully replacing one player at a time, Systematic Replacements aim to create squad variants and guarantee a comprehensive investigation of various combinations and their expected performance. By using machine learning to classify teams according to their performance and credit scores, K-Means Clustering finds the best team configurations using cluster analysis and data-driven insights. Each method provides unique advantages for both pre-match predictions and post-match evaluations, enhancing the decision-making process in fantasy cricket.

3.1. Random Sampling

Algorithmic Description for Fantasy Cricket Team Formation using Random Sampling Approach:

1. Fetch the pool of players with their attributes (e.g., performance score, cost, and role).
2. Set the number of teams to be created (team_count = 100).
3. For each team:
   a. Randomly select 11 players from the pool.
   b. Check if the selected team meets credit and role restrictions.
   c. If valid, add the team to the list of generated teams.
4. For each generated team:
   a. Calculate the predicted performance score based on historical data.
5. After the match:
   a. Retrieve actual performance data for each player.
   b. Calculate the actual performance score for each team.
6. Analyze and visualize the performance of all teams.

Figure 1 and Figure 2 shows the Distribution of Randomly Created 100 Fantasy Cricket Teams Prior to the Match and After the Match.

Journal of Theoretical and Applied Information Technology
31st January 2025. Vol.103. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.1.1. Pre-Match Analysis

Based on roles (WK, BAT, AR, and BOW) and historical performance data, players are selected. Every player has characteristics such as performance score and cost of credits.

- **Team Formation:** Using a random selection process, create 100 fantasy teams by choosing 11 players from the player pool. Make sure all teams adhere to credit limits and role restrictions.

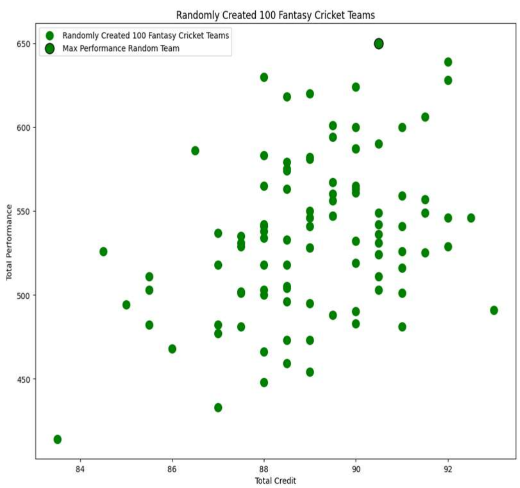

Figure 1: Distribution of Randomly Created 100 Fantasy Cricket Teams Prior to the Match.

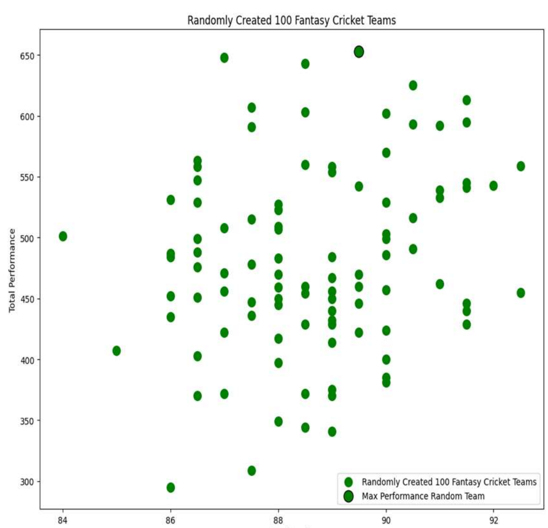

Figure 2: Distribution of Randomly Created 100 Fantasy Cricket Teams after the Match.

- **Performance Prediction:** Add together each player's score to determine each team's expected performance score. Use a score distribution visualization to find possible high-performing teams.

3.1.2. Post-Match Analysis

To determine each player's performance score, use real match data. Add together each team's performance points. Based on overall scores, compare and order the teams. To find the best-performing teams, visualize the score distribution. Examine participant input to make the best decisions going forward.

3.2. Systematic Replacements

Algorithmic Description for Fantasy Cricket Team Formation using Systematic Replacements Approach:

1. Create an initial valid team by selecting players from the pool.
2. Initialize an empty list to store team variations.
3. For each player in the initial team:
   a. For each player in the pool not already in the team:
   b. Replace the player in the initial team with the new player.
   c. Check if the new team meets credit and role restrictions.
   d. If valid, add the new team to the list of variations.
4. For each team variation:
   a. Calculate the predicted performance score based on historical data.
5. After the match:
   a. Retrieve actual performance data for each player.
   b. Calculate the actual performance score for each team.
6. Analyze and visualize the performance of all variations.

Figure 3 and Figure 4 shows the Distribution of Manually Created 100 Fantasy Cricket Teams Prior to the Match and After the Match.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.2.1. Pre-Match Analysis

Create the first two teams, players from team1 and players from team2. Choose players from these starting teams to form your fantasy squad. To create different squad configurations, replace players one at a time in a methodical manner. Make that every variation complies with job and credit requirements. Estimate the performance ratings for every variation. See how the scores of each variation are ordered.

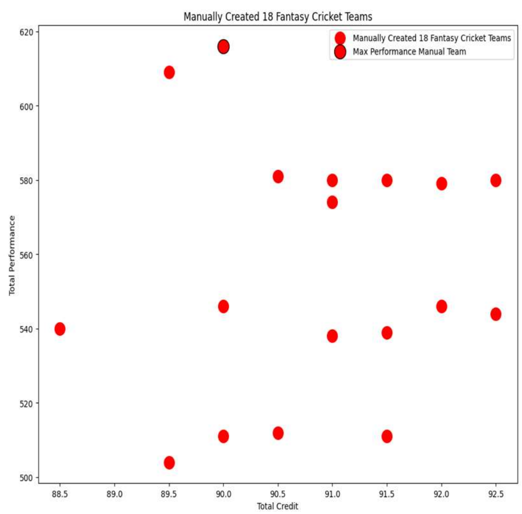

Figure 3: Distribution of Manually Created 18 Fantasy Cricket Teams prior to the Match.

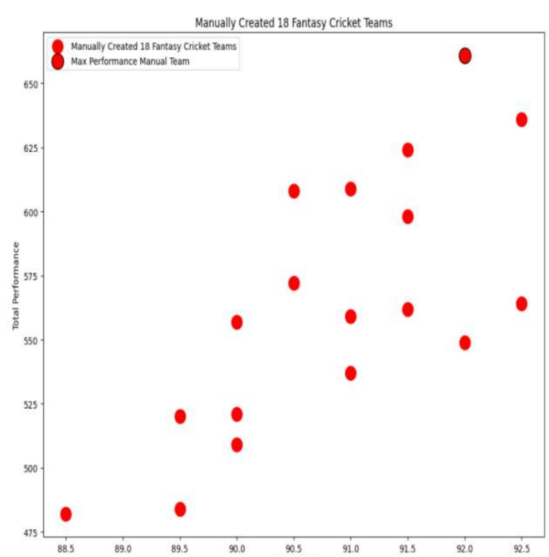

Figure 4: Distribution of Manually Created 18 Fantasy Cricket Teams after the Match.

3.2.2. Post-Match Analysis

Utilizing match data, determine the real performance scores. Add together all of the team variations' scores, then compare. Determine which team is performing the best. See performance scores graphically to see how various combinations affect things. Refine replacement techniques with insights.

3.3. K-Means Clustering

Algorithmic Description for Fantasy Cricket Team Formation using K-Means Clustering Approach:

1. Fetch the pool of players and prepare player data, including attributes like credits and historical performance.
2. Set the number of clusters (num_clusters = 5).
3. Perform K-Means clustering on the player data:
    a. Use performance and credit as clustering features.
    b. Group players into clusters.
4. For each cluster:
    a. Select an optimal team configuration that meets credit and role restrictions.
    b. Add the valid team to the list of optimal teams.
5. For each optimal team:
    a. Calculate the predicted performance score based on historical data.
6. After the match:
    b. Retrieve actual performance data for each player.
    c. Calculate the actual performance score for each team.
7. Analyze and visualize the clusters and performance of all optimal teams.

Figure 5 and Figure 6 shows the Distribution of K-Means clustering Fantasy Cricket Teams Prior to the Match and After the Match.

3.3.1. Pre-Match Analysis

Gather information about the players, such as their roles, credits, and performance history. Establish the parameters for the team's formation (total players, credit limit, role limits). Within the limitations, create a valid team configuration. Utilize credit and overall performance as clustering features. Teams should be grouped using K-Means according to credit and performance. Examine clusters to find the best team arrangements.

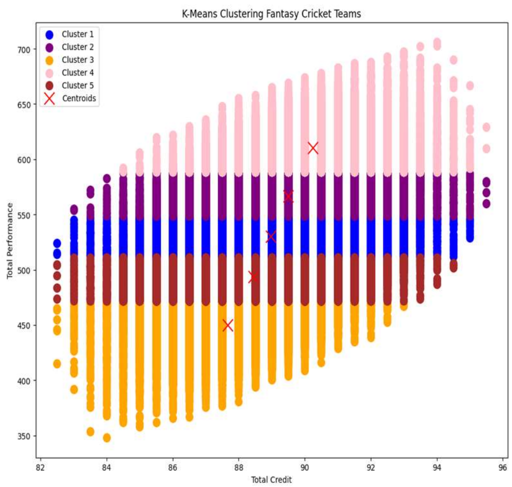

Figure 5: Distribution of K-Means Clustering Fantasy Cricket Teams prior to the Match.

3.2.2. Post-Match Analysis

Player performance scores are calculated using the results of the actual matches. Teams should be regrouped according to credit and real performance. To find the best teams, compare clusters. Highlight the best-performing configurations by visualizing centroids and clusters. Make use of the findings to guide future strategy and team decisions.

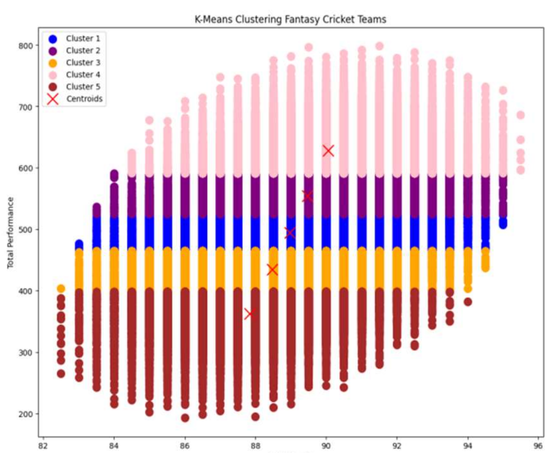

Figure 6: Distribution of K-Means Clustering Fantasy Cricket Teams after the Match.

3.4. Key Steps and Functions in All Approaches

1. Player Pool Initialization: Fetch player data, including performance and cost metrics.
2. Team Validation: Ensure team configurations comply with credit and role restrictions.
3. Performance Calculation:
   a. Predicted: Use historical data.
   b. Actual: Use real match data.
4. Analysis and Visualization: Compare team performance using statistical and graphical methods.

4. EXPERIMENTAL RESULTS

Performed a number of tests to see which of the three approaches—manual, random, and K-Means clustering—was better at creating the best fantasy cricket teams. Based on the combined performance score of the teams that were established, each method's effectiveness was evaluated.

4.1. Manually Produced Teams

We manually produced eighteen fantasy cricket teams using a meticulous selection process based on player roles, credits, and performance indicators. Computed and examined the teams' overall performance scores that were manually formed.

To evaluate the effectiveness of the three methods (manual, random, and K-Means clustering) in forming optimal fantasy cricket teams, we conducted a series of experiments. The performance of each method was measured based on the total performance score of the teams formed. Manually Created Teams: We manually created 18 fantasy cricket teams by carefully selecting players based on their roles, credits, and performance metrics. The total performance scores for these manually created teams were calculated and analyzed.

4.2. Teams Created at Random

Using a random selection process, we created 100 fantasy cricket teams by choosing 11 players at random from the available pool. To establish a standard for comparison, the performance scores of these randomly selected teams were calculated.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.3. K-Means Clustering

Based on each player's performance and credits, we grouped them using the K-Means clustering algorithm. Within each cluster, several team configurations were created to guarantee that the limitations on the total credits and role distribution were followed. Next, the performance ratings of these groups of teams were assessed, as shown in the Figure 5 and Figure 6.

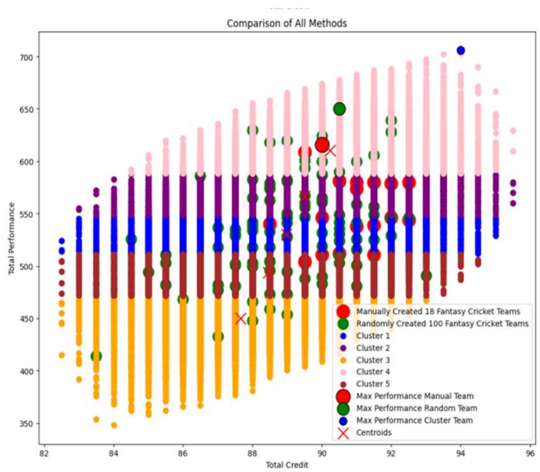

Figure 7: Distribution of User-Created, Random, and K-Means Clustering Fantasy Cricket Teams prior to the Match.

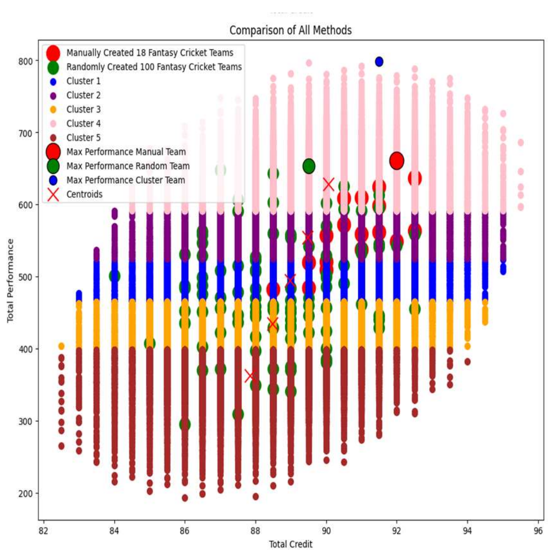

Figure 8: Distribution of User-Created, Random, and K-Means Clustering Fantasy Cricket Teams after the Match.

4.1. Performance Comparison

The maximum performance scores for each method were: Manual: 619, Random: 607, K-Means Clustering: 635. The mean performance scores for each method were calculated as follows in the equations 1, 2 and 3:

$$Manual\ Teams\ Mean = \frac{\sum_{i=1}^{n} x_i^{(manual)}}{n} \quad (1)$$

$$Random\ Teams\ Mean = \frac{\sum_{j=1}^{m} x_j^{(random)}}{m} \quad (2)$$

$$K-Means\ Clustering\ Mean = \frac{\sum_{l=1}^{k} x_l^{(kmeans)}}{k} \quad (3)$$

Where:

- **n** is the number of manual teams, and $\mathbf{x_i^{(manual)}}$ is the performance score of the **i-th** manual team.
- **m** is the number of random teams, and $\mathbf{x_j^{(random)}}$ is the performance score of the **j-th** random team.
- **k** is the number of clusters, and $\mathbf{x_\ell^{(kmeans)}}$ is the (average) performance score for the **ℓ-th** cluster found by k-means.

4.2. Error Calculation:

To calculate the error, we considered the deviation of each team's performance from the highest achievable performance within the dataset. The root mean square error (RMSE) was used as a metric to quantify this deviation as shown in the equation 4:

$$\boldsymbol{RMSE = \sqrt{\frac{\sum_{i=1}^{N}(perfor - max\ perfor)^2}{N}}} \quad (4)$$

Where,

- **perfor$_i$**: The performance score of the **i-th** team.
- **max perfor**: The highest performance score achieved (635 in this case).
- **N**: The total number of teams evaluated.

4.3. Results

These RMSE values indicate the average deviation of the team performance scores from the highest performance score, with lower RMSE values reflecting more effective optimization of team performance as shown in the Table 1.

Table 1: RMSE values of the methods.

Method	Pre Match RMSE	Pre Match Mean	Post Match RMSE	Post Match Mean
Manual	69.32	555	108.97	564
Random	93.65	539.21	194.78	481.76
Clustered	98.70	529.58	184.32	493.96

5. CONCLUSION

This study examined three different approaches to building fantasy cricket teams: K-Means clustering, random generation, and manually constructed teams. Teams that were manually assembled showed a well-balanced configuration of credits and performance, reaching high maximum performance at the expense of being time-consuming and having a narrow scope. These teams made use of human skills and strategic thought. Though they frequently produced subpar teams and lacked strategic input, randomly generated teams, on the other hand, offered a wide range of performance and credit utilization, exhibiting varied options and occasionally providing competitive outcomes.

Similar performance and credit usage teams were efficiently grouped by the K-Means clustering algorithm, which also systematically optimized configurations and often outperformed the best manually and randomly constructed teams. This machine learning method strikes a balance between optimization and exploration, which makes it an effective tool for team building. In conclusion, K-Means clustering emerges as the most successful technique, continuously improving team performance and providing a data-driven method to improve decision-making in fantasy sports team construction, even if all three approaches may create competitive fantasy cricket teams.

REFERENCES:

[1]. 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* 325, 289–317 (2023). https://doi.org/10.1007/s10479-022-04726-z

[2]. Gerber, H., & Sharp, G. D. (2006). Selecting a limited overs cricket squad using an integer programming model. *South African Journal for Research in Sport, Physical Education and Recreation*,*28*(2). https://doi.org/10.4314/sajrs.v28i2.25945

[3]. Dey, P. K., Ghosh, D. N., & Mondal, A. C. (2011). A MCDM approach for evaluating bowler's performance in IPL. *Journal of Emerging Trends in Computing and Information Sciences, 11*, 563–573. https://www.academia.edu/download/7930959/vol2no11_2.pdf

[4]. Sachin Kumar S, Prithvi HV, C Nandini, "A data science approach to predict the winning fantasy cricket team Dream 11 Fantasy Sports", https://doi.org/10.48550/arXiv.2209.06999

[5]. S. Chauhan, R. Kumar and B. Kumar, "Machine Learning Approaches to Predict the Teams for Fantasy Leagues," *2023 International Conference on Recent Advances in Information Technology for Sustainable Development (ICRAIS)*, Manipal, India, 2023, pp.54-58. https://ieeexplore.ieee.org/document/10367068

[6]. Vetukuri, V.S., Sethi, N. & Rajender, R. Generic model for automated player selection for cricket teams using recurrent neural networks. *Evol. Intel.* **14**, 971–978 (2021). https://doi.org/10.1007/s12065-020-00488-4

[7]. Y. Kumar, H. Sharma and R. Pal, "Popularity Measuring and Prediction Mining of IPL Team Using Machine Learning," *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India,2021,pp.1-5. https://ieeexplore.ieee.org/document/9596405

[8]. Bedekar, M. V., Pande, M., Vishwarupe, V., Joshi, P. M., Pawar, V. & Shingote, P. (2022). Data Analytics in the Game of Cricket: A Novel Paradigm. Procedia Computer Science, 204:937–944. https://doi.org/10.1016/j.procs.2022.08.114

[9]. Farquhar, L. K., & Meeds, R. (2007). Types of Fantasy Sports Users and Their Motivations. *Journal of Computer-Mediated Communication*, *12*(4), 1208–1228. https://doi.org/10.1111/j.1083-6101.2007.00370.x

[10]. Drayer, J., Dwyer, B., & Shapiro, S. L. (2013). Examining the impact of league entry fees on online fantasy sport participation and league consumption. European Sport Management Quarterly, 13(3), 339–357.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

https://doi.org/10.1080/16184742.2013.783605

[11]. Dwyer, B., Shapiro, S.L. & Drayer, J. Daily Fantasy Football and Self-Reported Problem Behavior in the United States. *J Gambl Stud* **34**, 689–707 (2018). https://doi.org/10.1007/s10899-017-9720-4

[12]. Dwyer, B., Shapiro, S. L., & Drayer, J. (2024). Fan conflict or complement? Exploring the intersection of sports betting and team loyalty. International Journal of Sports Marketing and Sponsorship,25(3),483–498. https://doi.org/10.1108/IJSMS-07-2023-0138

[13]. Kissane, R. J., & Winslow, S. (2016). Bonding and Abandoning: Gender, Social Interaction, and Relationships in Fantasy Sports. *Social Currents*, *3*(3), 256-272. https://doi.org/10.1177/2329496515616821

[14]. Eric Fisher, 2008, "Study: Fantasy players spendbig". https://www.sportsbusinessjournal.com/Journal/Issues/2008/11/17/This-Weeks-News/Study-Fantasy-Players-Spend-Big.aspx

[15]. Kumarasiri, I., & Perera, S. (2017). Optimal One Day International Cricket Team Selection by Genetic Algorithm. *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, *36*(4), 213–221. Retrieved from https://gssrr.org/index.php/JournalOfBasicAndApplied/article/view/8097