Behaviour of Players on IPL Based on Fuzzy C Means

Sandeep S. Jain, Riya Gupta, Chetika Tiwari, Navnoor Kaur

Abstract: Clustering algorithms are being widely used in the field of data mining in order to accumulate similar data in the form of clusters. Indian Premiere League(IPL) is one of the most famous cricket leagues around the globe. In this paper, the dataset of IPL is used to cluster the players on the basis of various attributes. The authors ought to analyze both batsmen and bowlers in various clusters with the help of Fuzzy-c-means. The algorithm has been implemented to group the players in different clusters based on their performance in the IPL season of 2018. The pros and cons of the algorithm are also discussed and finally the experimental results are shown to highlight two main clusters i.e. above average and below average. The present work simulates the algorithm to distinguish only overseas player. In future this work can be extended for every player and form a recommendation model to identify best player or form best team.

Index Terms – Clustering algorithm, Fuzzy-C-Means, Clusters, IPL Dataset.

I. INTRODUCTION

Cricket is the second most famous sport in the world with over 2.5 billion fans around the globe. It has a global presence especially in the countries like India, Pakistan, West Indies, Sri Lanka and Australia, who were under the British Rule in the past [1]. As the game requires sportsmanship, commitment and discipline it's also called as a gentleman's game [2]. There is an annual cricket tournament which takes place in India known as the Indian Premiere League(IPL) which was first played in the year 2008 [3]. The players comprise of both domestic as well as international players who are selected in an auction [1]. IPL is hosted by the Board of Control for Cricket in India (BCCI), the governing body for cricket in India at both domestic as well as International levels [3]. IPL has a huge fan following and attracts huge sums of money for all the players, the governing body and the franchises [2].

Clustering is a vital non supervised learning tool and is applied in computer image programming tasks, e.g., image segmentation. There are many types of clustering algorithm such as K-Means, DBSCAN (Density Based Spatial Clustering of Application with Noise)[4].In the clustering, there are two sets as labelled set and candidate set, which assign pixels. Basically, there are three types of pixels like SLICERS, SEEDS etc.[5].

There are many different applications based on clustering which could be useful to approximate the value to detecting the clusters [6].There are many different applications based on clustering which could be useful to approximate the value to detecting the clusters[6]. Mostly, clustering can be divided into two parts like partitioning and hierarchal clustering [7].Image segmentation is the most vital and hard problems in many applications, medical processing, robot vision etc. [8]. Clustering algorithms is the vital role to discovering useful knowledge from expandable databases [9]. It is also a mathematical tool that attempts to discover structures [9].

A. Materials and Methods

The process of clustering usually refers to the technique of investigating a number of patterns and organising them into different clusters so that patterns having resemblance are allotted to a single cluster [7].Fuzzy c-means is the one of the algorithms used in clustering which enables clustering procedure to absorb the more actual data and image information [8].It was firstly introduced by Zadeh in 1965.Moreover,its application was initiated by Zadeh, Ruspini and Bellman. Hoppner et al also proposed towards examining the fuzzy c-means [7].

Fuzzy c-means is a methodical clustering function which is affected by the feature weight. So, for the proper functioning of this algorithm feature weights should be adequately chosen [9].

Basically, fuzzy clustering problems are differentiated into three divisions based on:

 (i) fuzzy rule learning

(ii) fuzzy relation

(iii) Maximum efficiency of an indented function [7].

It is perfect unchecked method for the elevation of the models and the inspection of the given data [9]. It also contains 2-D scattered regular or irregular data patterns [10]. The basic idea in fuzzy clustering is that it gives best results where there is overlapping of the data patterns [10].

B. Working

Steps to perform fuzzy c-means algorithm are as follows:

Let P = {p1, p2, p3 ..., p_n} be the set of data points and

C = {c1, c2, c3 ..., c_x} be the set of centers.

1) Randomly select 'x' cluster centers.

2) Calculate the fuzzy membership u_{ab} using:

$$u_{ab} = 1 / \sum_{j=1}^{x} \left(\frac{d_{ab}}{d_{aj}}\right)^{\left(\frac{2}{v}-1\right)} \qquad \text{Eq.1}$$

3) Compute the fuzzy centres 'c_b' using:

$$c_b = \left(\sum_{a=1}^{n} (u_{(ab)}^{v} p_a\right) / \sum_{a=1}^{n} (u_{(ab)})^{v} \qquad \text{Eq.2}$$

for b=1,2,3 ……x.

4) Repeat step (2) and (3) until the minimum 'K' value is achieved or $\|M^{j+1} - M^j\| < \alpha$.

$$K_v = \sum_{a=1}^{n} \sum_{b=1}^{t} (u_{ab})^{v} \|P_a - C_b\|^2 \qquad , \text{eq.3}$$

where,

'j' is the iteration step.

'α' is the termination criterion between [0, 1].

M = (u_{ab})n *x is the fuzzy membership matrix.

'K' is the objective function.

$\|P_a - C_b\|^2$ represents Euclidean distance

between ath and bth cluster center.

C. Advantages

(1) The advantage of fuzzy c-means algorithms, is in the application of digital image processing and image segmentation such as segmentation of the brain MR images [7].

(2) The main benefit of fuzzy c-means is that it deals with the ambivalent data consisting clusters of different shapes such as ellipsoidal, linear, spherical or other shapes [11].

(3) Fuzzy c-means has its productive application in various fields such as segmenting video, medical sector, intrusion diagnosis, etc.[7].

(4) Learning time is the shortest and its error rates is less than other methods.

D. Disadvantages

(1) The presence of noise influences the fuzzy c-means algorithm [9].

(2) Distance metric selection also fragile this algorithm [9].

(3) Time taken to apply this method depends upon size of image. Therefore, complex images require larger amount of time [11].

(4) Based on findings of various researches, this method does not have a persistent supremacy in the varied structures of data [10].

II. METHODOLOGY AND EVALUATION

Clustering algorithms can be used in the field of Sports as they can be used to monitor the progress of the players. With the help of algorithms such as Fuzzy-C-means, K means etc. separate clusters are formed to differentiate sportsmen on the basis of their performance. The better

players are kept in a cluster whereas others are kept at a far distance from them.

The dataset used in this paper depicts the performance of various players who took part in the Indian Premiere League(IPL) in the year 2018. This league is held annually and provides a prodigious opportunity for the cricketers around the globe to showcase their skills here in India. The dataset comprises of around 144 players who play different roles on the field of cricket. These players can be classified as Batsmen, Bowlers and All-Rounders.

Firstly, both the batsmen and bowlers are classified without using Matches as any dependent attribute and then we took matches in the consideration in order to explore if the experience of a player counts in his performance or not.

Batsmen and Bowlers are generally monitored on the basis of various categories mentioned below. With the help of Fuzzy-C-Means algorithm we are able to cluster the batsmen on basis of the following attributes:

1) **Total Runs (TR):** The total runs scored by any player during the whole league.

2) **Highest Score (HS):** Highest score of a batsman in a single inning.

3) **Average Score (AvgS):** The average sum total of a batsmen in the overall league.

4) **Strike Rate (SR):** Average number of runs scored per 100 balls faced by any batsmen.

5) **Matches (Mat):** Total number of matches played by the batsman.

Similarly, bowlers can also be classified on the basis of the following attributes:

1) **Overs (Ov):** The total number of overs bowled by a bowler.

2) **Runs (R):** Total runs conceded by a bowler during the entire league.

3) **Wickets (Wkt):** Total number of wickets taken by a bowler.

4) **Average (Avg):** Average runs yielded by a bowler during the complete season.

5) **Economy (E):**Average runs yielded for each over bowled.

6) **Strike Rate (SR):** Measurement of a bowler's average balls bowled for each wicket taken.

7) **Matches (Mat):** Total number of matches played by thebowler.

First of all, in this article batsman are clustered on the basis of attributes/features explained above. Only one attribute has not been kept into consideration i.e. "Matches.X".The aim was to check whether the experience of a player counts in his performance or not.

151

As quoted by the author in his paper, both batsmen and bowlers were classified on the basis of the following limited attributes:

1. Average

2. Strike Rate

3. Economy

The players were studied on the basis of the countries of their origin [12]. Whereas, we tried to group the players on the basis of their performance.Hence, we are able to predict that whether any IPL team should look to buy players of that specific country or not. Not only this, the owners would also be at benefit as they would only have to use their money to buy those players who are in the better cluster.

Table 1 Glimpse of the dataset of the batsmen

PLAYERS	TR.X	HS	Avg. S.X	SR.X
AB de Villiers	480	90	53.33	174.54
AjinkyaRahane	370	65	28.46	118.21
Chris Gayle	368	104	40.88	146.03
GautamGambhir	85	55	17	96.59
MS Dhoni	455	79	75.83	150.66

Table 1 depicts the data of the batsmen where the number of matches playedare not taken into consideration. Similarly, table 2 shows the data for the batsmen where the number of matches played are considered in the scenario.

Table 2Dataset of batsmen where matches are included

PLAYERS	Matches. X	TR. X	HS	Avg.S. X	SR.X
AB de Villiers	12	480	90	53.33	174.54
AjinkyaRahane	15	370	65	28.46	118.21
Chris Gayle	11	368	104	40.88	146.03
GautamGambhir	6	85	55	17	96.59
MS Dhoni	16	455	79	75.83	150.66

Secondly,similar to the previous experiment, the same operation was performed on the bowler's dataset. Firstly, with considering the number of matches played by the bowler and then taking the scenario where the experience of the player was not considered.

Table 3 Describing bowler's performance in the IPL 2018 on the basis of total matches played

PLAYERS	A Russel	BKumar	RAshwin	R Jadeja
Mat.Y	16	12	14	16
Ov	37.5	46.1	50.4	41
R.Y	355	354	410	303
Wkt	13	9	10	11
Avg.Y	27.3	39.33	41	27.54
E	9.38	7.66	8.09	7.39
SR.Y	17.46	30.77	30.4	22.36

Table 4 depicts the data of the bowlers where the number of matches played are not taken into consideration

Player	ARussel	B Kumar	R Ashwin	R Jadeja
Ov	37.5	46.1	50.4	41
R.Y	355	354	410	303
Wkt	13	9	10	11
Avg.Y	27.3	39.33	41	27.54
E	9.38	7.66	8.09	7.39
SR.Y	17.46	30.77	30.4	22.36

Now to move further, the number of clusters were increased from two to four. The number of clusters were varied from two to four but the results obtained were always same. The change in number of clusters always resulted in formation of two clusters, as the other two clusters remained empty. Hence, the data was grouped in two clusters, which can be said as- above average and below average as described in the graphs.

III. RESULTS AND DISCUSSIONS

Fig.1 shows the result obtained from our clustering analysis. Here the players who are divided on the basis of their country of origin are plotted on x-axis whereas total number of players are represented by the y axis. Hence this graph shows two different clusters where all those batsmen who are below average are represented by Cluster 1 and the ones above average by Cluster 2. Not only this, the graph helps to find out which foreign country produces better batsmen as compared to others.

When we took number of matches into consideration, the expected results were out, as shown in Fig.2. The experience of a player has played an important factor here as several players which were earlier found in Cluster1 have shifted to Cluster2.

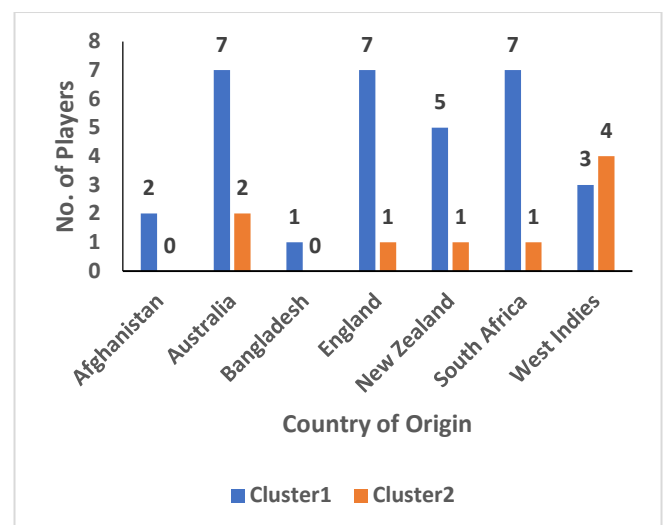*Fig.1 Players being divided in their respective countries and then classified on the basis of their cluster.*

Behaviour of Players on IPL Based on Fuzzy C Means

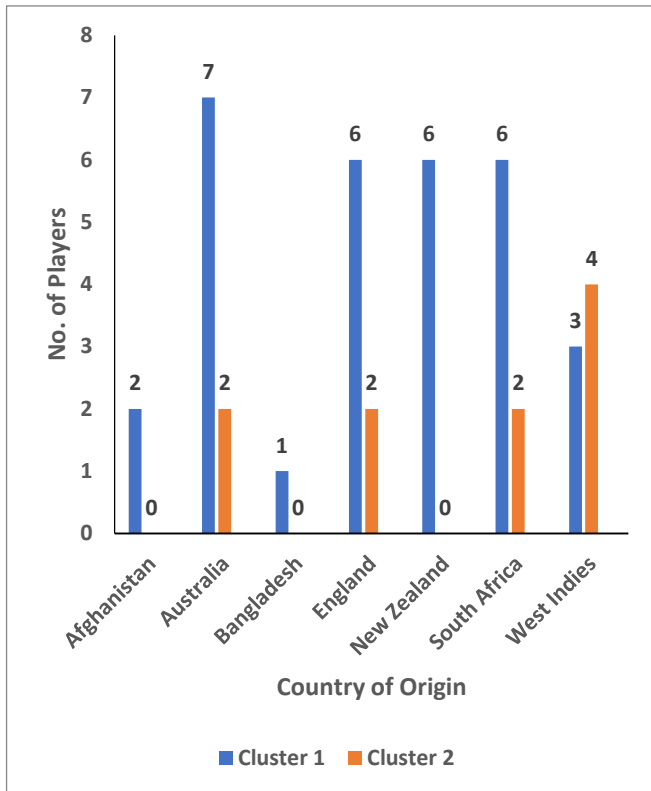*Fig.2 Players being clustered when their experience was included in the dataset.*

Fig. 3 Bowlers being clustered when the total number of matches were taken into consideration in the dataset.
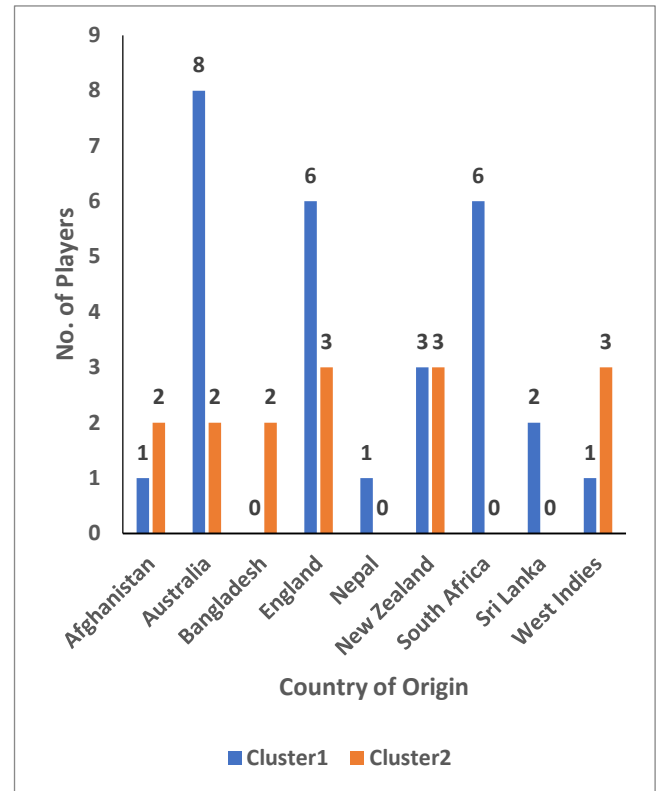

Table 3 describesbowler's performance in the IPL 2018 on the basis of total matches played

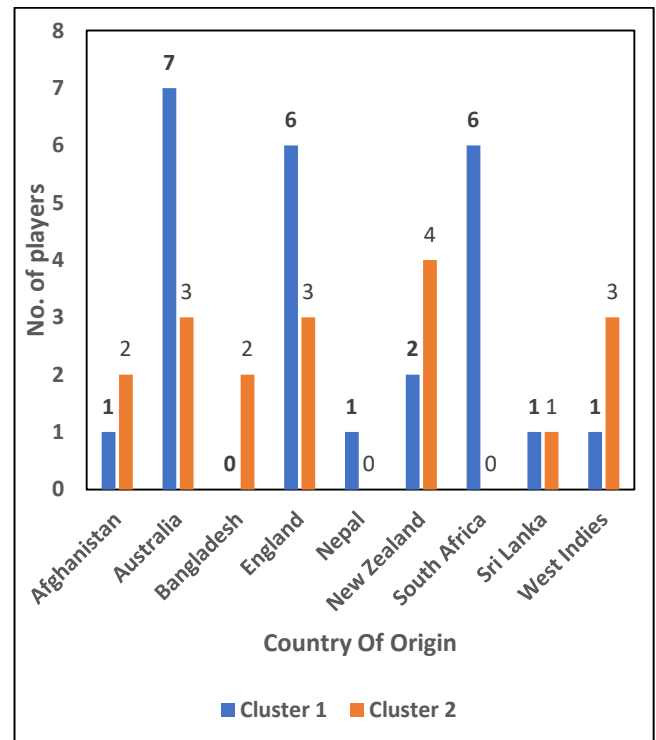| PLAYER	A Russel	BKumar	RAshwin	R Jadeja
Mat.Y	16	12	14	16
Ov	37.5	46.1	50.4	41
R.Y	355	354	410	303
Wkt	13	9	10	11
Avg.Y	27.3	39.33	41	27.54
E	9.38	7.66	8.09	7.39
SR.Y	17.46	30.77	30.4	22.36

Table 4 depicts the data of the bowlers where the number of matches played are not taken into consideration

Player	ARussel	B Kumar	R Ashwin	R Jadeja
Ov	37.5	46.1	50.4	41
R.Y	355	354	410	303
Wkt	13	9	10	11
Avg.Y	27.3	39.33	41	27.54
E	9.38	7.66	8.09	7.39
SR.Y	17.46	30.77	30.4	22.36

Fig.4 The results of the bowlers where the number of matches played by the bowlers were not considered

After comparing both Fig.3 and Fig.4 we got to know that the experience of the player counts while analyzing a bowler's performance.

Now,the number of clusters were varied from two to four but the results obtained always formed only two clusters, as the other two clusters remained empty. Hence, the data was grouped in two clusters, which can be said as- above average and below average as shown in tables.

Table 5 represents the dataset where Noise Clustering(NC) level is 4.

PLAYER	Mat. X	RUNS .X	HS	AVG .X	SR.X	CLUSTER
A. Rahane	15	370	65	28-46	118	4
Chris Gayle	11	368	104	40.88	146.03	4
G.Gambhir	6	85	55	17	96.59	1
MS Dhoni	16	455	79	75.83	150.66	4
R.Jadeja	16	89	27	17.8	120.27	1

Table 5 shows that despite keeping the Noise Clustering(NC)level at 4, still we were able to classify our data into two clusters only. Hence the method of NC=3 and NC=4 has to be terminated.

IV. CONCLUSION

Fuzzy-c means algorithm turns out to be a much better algorithm as compared to k-means especially when the dataset has outliers or noise. From the results, we were able to conclude that the players present in Cluster2 are better as compared to those in Cluster1. In terms of overseas batsmen, West Indies is able to produce better quality of batsmen as compared to any other foreign country. Whereas, in terms of overseas bowler's, Bangladesh has only 2 bowlers participating in the IPL 2018, but both are exceptionally well as compared to those of any other country. We aim to normalize every player with a score and form a recommendation system for both batsman and bowler in the near future.

ACKNOWLEDGMENT

REFERENCES

[1] Swartz, T. B., Research directions in cricket. Handbook of Statistical Methods and Analyses in Sports. Chapman & Hall/CRC Handbooks of Modern Statistical Methods: Boca Raton,2016, FL, 272.

[2] Shah, S., Hazarika, P. J., & Hazarika, J.,"A Study on Performance of Cricket Players using Factor Analysis Approach."International Journal of Advanced Research in Computer Science, 8(3), 2017.

[3] Mittal, A., Manavalan, A., "The IPL Model: Sports Marketing and Product Placement Sponsorship"International Journal of Humanities and Social Science Invention, 6(2), 2017.

[4] Hou, J., Gao, H., & Li, X.,"DSets-DBSCAN: a parameter-free clustering algorithm." IEEE Transactions on Image Processing, 25(7), 2016, pp. 3182-3193.

[5] Shen, J., Hao, X., Liang, Z., Liu, Y., Wang, W., & Shao, L,"Real-time superpixel segmentation by DBSCAN clustering algorithm." IEEE Transactions on Image Processing, 25(12), 2016, pp. 5933-5942.

[6] Ienco, D., &Bordogna, G. ,Fuzzy extensions of the DBScan clustering algorithm. Soft Computing, 22(5), 2018, pp. 1719-1730.

[7] Nayak, J., Naik, B., &Behera, H. S. ,"uzzy C-means (FCM) clustering algorithm: a decade review from 2000 to 2014." Computational intelligence in data mining-volume 2, 2015, pp. 133-149.

[8] Zheng, Y., Jeon, B., Xu, D., Wu, Q. M., & Zhang, H. ,"Image segmentation by generalized hierarchical fuzzy C-means algorithm." Journal of Intelligent & Fuzzy Systems, 28(2), 2015, pp. 961-973.

[9] Pourjabari, A. J., &Seyedzadegan, M.," An improved method of fuzzy c-means clustering by using feature selection and weighting." International Journal of Computer Science and Network Security (IJCSNS), 16(10), 2016, pp.64.

[10] Cebeci, Z., &Yildiz, F.,"Comparison of K-means and Fuzzy C-means algorithms on different cluster structures." AGRÁRINFORMATIKA/JOURNAL OF AGRICULTURAL INFORMATICS, 6(3), 2015, pp. 13-23.

[11] Ludwig, S. A.,"MapReduce-based fuzzy c-means clustering algorithm: implementation and scalability." International journal of machine learning and cybernetics, 6(6), 2015, pp. 923-934.

[12] Chirag Goyal, "IPL-2017 Cross Country Cluster Analysis."International Journal of Computer Science Trends and Technology (IJCST)5(4),2017, pp.117-124.
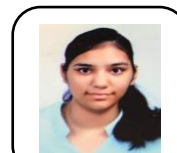
AUTHORS PROFILE

Sandeep S. Jain, pursuing B.Tech. in Computer Science at ChitkaraUniversity,Pajpura,Punjab. Fixed delegate of Global goals Model United Nation(GGMUN) 2019 held in Malaysia. Former member of Toastamsters.

RiyaGupta is pursuingB.Tech in Computer Science at ChitkaraUniversity,Rajpura, Punjab.

Chetika Tiwariispursuing B.Tech in Computer Science at ChitkaraUniversity,Rajpura, Punjab.

Navnoor Kaur ispursuing B.Tech in Computer Science at ChitkaraUniversity,Rajpura, Punjab.

Retrieval Number: I10240789S19/19©BEIESP
DOI: 10.35940/ijitee.I1024.0789S19

154

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication