https://doi.org/10.48047/AFJBS.6.13.2024. 3312-3329

AFJBS

African Journal of Biological Sciences

Clustering of Agriculture Data through Fuzzy C-Means Technique

BanshidharChoudhary[a] andVipinSaxena[b]
Department of Computer Science
BabasahebBhimraoAmbedkar University
Lucknow (UP), 226025, India
[a]banshidharphd@gmail.com, [b]profvipinsaxena@gmail.com

Article History

Volume 6, Issue 13, 2024

Received: 18June 2024

Accepted: 02July 2024

doi:***10.48047/AFJBS.6.13.2024.*** 3312-3329

ABSTRACT

As the data is growing over the cloud servers and user's are downloading the data through the hand-held devices attached via cloud srvers, it is necessary to extract the desired information in the less time. In the present paper, an efficient fuzzy based C-Means clustering of data is applied over the real agriculture database related to the wheat and rice grains for optimizing the time of extracting the desired information by the users. It is observed that the fuzzy C-Mean technique applied over the selected database provides the optimize results in comparison of K-Means and computed results in terms of efficiency using python programming and R language are depicted in the form of tables and figures.

Keywords: Cloud Servers, Agriculture Database, Wheat and Rice Grains, Fuzzy C-means and Efficiency

1. INTRODUCTION

Agriculture is a major contributor to the economy of many countries among the world. The agriculture sector is generating a large volume of data related to crop yield, soil composition, weather condition and many other factors that effects agriculture production. The data can provide valuable insight that may be able to be used to improve crop yield and make more information decision in the agriculture industry. Clustering is a popular technique in data mining that involves similar objects or data points together. Clustering can help to identify the pattern and relationship within large datasets, making it useful tool for analysing agriculture data .Fuzzy C-Means (FCM) is a clustering algorithm that allows data points belong to multiple clusters, which makes it particularly useful for data sets that have overlapping patterns or datapointsthat are difficult to assign to a single cluster. In the recent years, FCM has been applied to various agriculture datasets to identify relationships between different factors affecting crop yield.

The present work is focusing on the use of FCM clustering technique for analysing agriculture data which will discuss for the processing of data prepration, clustering and evaluation of results. The goal of the present work is to demonstrate the effectivness of FCM in clustering for agriculture data and to provide insight that can improve the agriculture production.

2. LITERATURE REVIEW

Cordoba et al. [1] have introduced a novel technique known as KM-sPC . In the subfield management area for defining within-field management classes. The approach coupled fuzzy K-Means cluster analysis with MULTISPATI-PCA to efficiently take into account spatial auto correlation in soil variables. The study's findings showed that KM-sPC outo performed non-spatial clustering strategies in improving within-field management class development, which is characterized by contiguous zoning and significant yield disparities between the demarcated classes.Bansod and pandey [2] have analysed Principal Component Analysis (PCA) and Fuzzy C-Means (FCM) clustering algorithms were used to look for site-specific spatial variability in apparent Electrical conductivity (EC), crop production, and several soil characteristics. The analysis included looking into how these variables related to one another. Using PCA and FCM clustering algorithms, the findings provided insights into the regional patterns of ECa, crop production, and soil characteristics.. Speranza et al. [3] have developed data mining techniques and a computer application that can generate maps of management zones and yield regions which were used to offer a cluster-based methodology for defining management zones in precision agriculture. Chang et al. [4] have studied to define Management Zones (MZ’s) for variable–rate fertilizer applications in tobacco-planting fields using an active canopy sensor to collect Normalized Difference Vegetative Index (NDVI) data at five growth stages. Patel and Patel [5] have summarized of the many data mining techniques used to model agricultural data, including fuzzy C-Means, neural networks, support vector machines, bi-clustering, K-nearest neighbor, and naive Bayes classifiers. The survey highlighted the methodologies' applicability in predictive modeling within the agricultural sector, where the usefulness depends on the particular agricultural data and problems at hand.Singhet al. [6] discussed the use of clustering techniques in precision agriculture and the advantages and disadvantages. It highlights the importance of applying technology and data analysis in maximizing agriculture production capacity. In the year 2015, Bazzi et al. [7] have identified soil and altimetry characteristics that influenced pear orchard yield and generate Management Zones (MZ’s) using Fuzzy C-Means algorithm. Two approaches were used i.e. chemical attributes and textural and altitude attributes. Rodrigues and Cora [8] have usedFuzzy C-Means clustering technique to identify management zones while taking both the geographical and temporal variations in soil properties and corn production into account. The study found that there were temporal fluctuations in the optimal count of management zones, highlighting the need of taking into account the temporal dynamics of crop output and soil properties for accurate delineation of management zones. Macroset al. [9] have determined the spatial interrelationships of soil physical properties,in a guava field in the semi-arid region of Brazil, research. In addition, fuzzy C-Means clustering analysis was used in the study to define management zones based on the discovered geographical connections. Another resource for understanding management classes in their context is the cited study. Speranzaet al. [10] have scussefully built a link between vegetation indices obtained from an RGB camera placed on an Unmanned Aerial Vehicle(UAV) and actual data to evaluate the nitrogen condition of wheat fields. The results showed a strong association between the

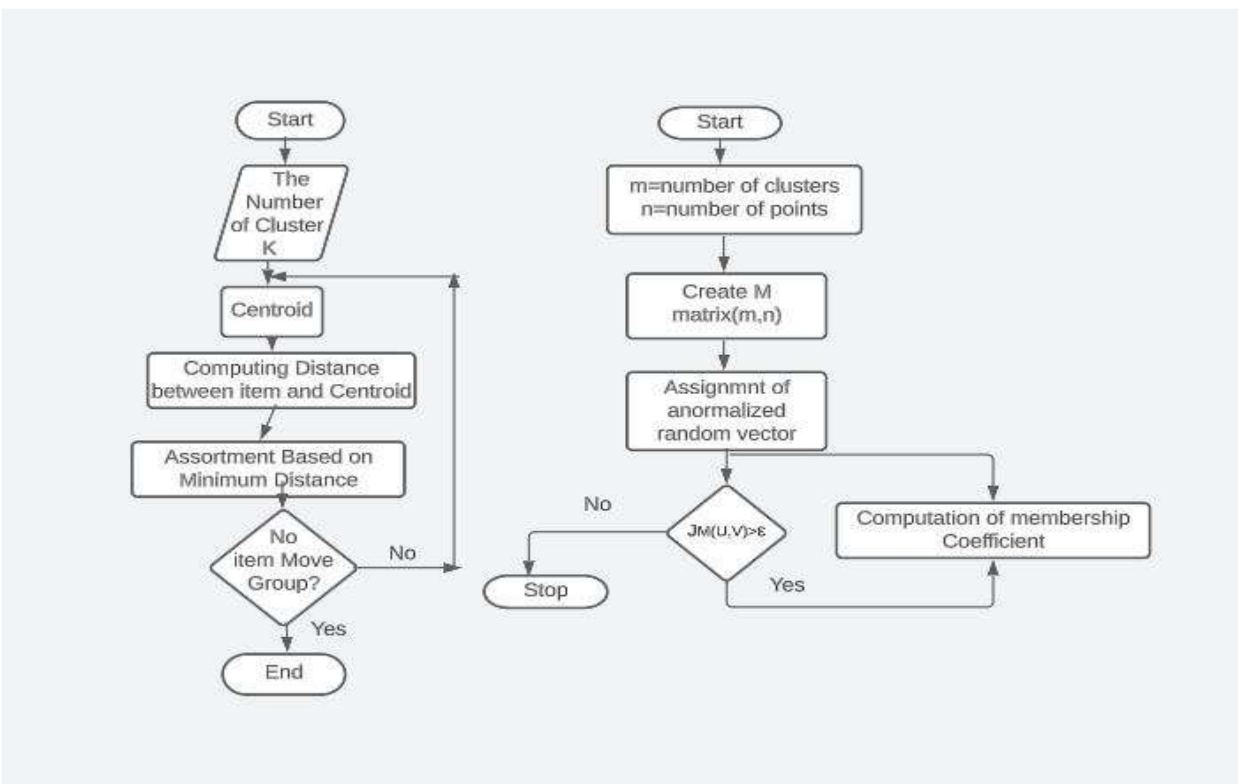
Normalized Green-Red Difference Index(NGRDI) , the Normalized Difference Vegetation Index(NDVI), and the values of SPAD and NDVI. This robust link remained for both varieties of wheat under investigation , confirming the usefulness of NGRDI as a trustworthy indicator for the purposes of the study. Gavioliet al. [11] have studied the assessed of three variables selection techniques and proposed a new approach, MPCA-SC, for defining management zones using the Fuzzy C-Means clustering algorithm. Ghoshet al. [12] have focused of the investigation on the use of complete wave form characteristics from the Ice, Cloud and Elevation Satellite/Geoscience Laser Altimeter System (GLAS) to categorize land cover in the western part of Doon Valley, Uttarakhand, India. The K-Means clustering methodology outperformed the other investigated approaches, with an overall classification accuracy of 89.41%. The three defined classes included a forest, a mango orchard, and a more general category that included other land uses like settlements, dry riverbeds, agricultural land, barren or fallow land, etc.Tamiminiaet al. [13]have developed a better multi-temporal PolSAR-based kernel-based C-Means clustering algorithm specifically for crop mapping. This strategy used the Particle Swarm Optimization (PSO) algorithm and polarimetric characteristics to optimize the procedure. Sindhuet al. [14] discussed the application of data mining techniques in agriculture for managing and processing large amounts of data to increase efficiency and productivity. Various techniques of data mining including fuzzy C-Means are described. The use of data mining can help to extract useful information from raw data and address various agriculture problems. Abdullahi1 et al.[15] have researched medicinal properties of vernoniaamygdalina, also known as bitter leaf,revealing its potential as an antidiabetic, antibacterial, antimalarial,antioxidant and cytotoxic agent compound such as steroid glucosides. Mohajer and Salehi [16] have determined and contrast how well the American Soil Taxonomy and World Reference Base soil categorization systems performed in describing the presence of heavy metal contamination in the Lenjant area of Isfahan, Iran. The inquiry was sparked by improvements in classification methods that took into account our growing understanding of human influence on soil formation.Camiciaet al. [17] studied and examined the variable rate seeding for soybean cultivation in two management zones, with productivity measured and mapped using harvest monitoring and inverse square distance interpolation. The result suggests a recommended seeding density of 214,000 plants per hectare in both zones based on economic analysis. Heupel et al.[18] have developed a unique method, using phenological growth and reflectance properties in multitemporal satellite pictures, to identify different crop kinds without the use of training data. The system, which combined fuzzy C-Means clustering and knowledgeable assessments of plant traits, produced an impressive overall accuracy of 89.49% over the 2015 growing season at a test site in Northern Germany. Cilliset al. [19] studied the use of precision agriculture principal to identify and characterized spatial variability in soil fertility and productive potential in field of Venice Lagoon. Rajput and Kumaravelu [20] have presentedan affordable clustering methodto achieve wireless sensor networks that are both sustainable and energy-efficient, specifically designed for agricultural monitoring in precision agriculture, an affordable clustering method was presented.Zahoet al. [21] have proposed a unique decision-making method, in order to improve precision fertilization methods, that combined Fuzzy C-Means (FCM) clustering with a Radial Basis Function (RBF) neural network fusion algorithm. The method significantly reduced computing time while improving forecast accuracy. Yassine et al. [22] have optimized wireless sensor networks in the context of agriculture, a clustering approach was devised. This strategy focused on reducing network energy usage while also optimizing operations at the network and data link levels.Pavan et al. [23] have discovered multi-attribute relationships including water quality physical characteristics in the context of agricultural crop yield production environments, a novel approach leveraging Pareto Optimal based Fuzzy C-Means clustering. Brubecket al.[24] have presented a novel approach to assessing biomass variability and creating management zones in a crop field using K-Means and fuzzy C-Means clustering algorithm. Liu et al. [25] have managed massive data from agricultural image processing, a novel parallel Fuzzy C-Means (FCM) segmentation algorithm based on Apache Spark . With an average speed of 12.54 on ten processing nodes, this method performed remarkably results. In terms of performance improvement, it outscored the Hadoop-based strategy by 128%. Bai et al. [26] have propossed a methodology for evaluating farmer creditworthiness using fuzzy rough-set theory and fuzzy C-Means clustering.Gavioliet al. [27] have proposed twenty methods for defining Management Zones (MZs) inside agricultural areas were subjected to a thorough review. Over ten approaches were included in this study that had not before been investigated for this purpose in published literature. The results showed that McQuality's Method and Fanny stood out as the best choices, producing internally coherent classes and significantly reducing yield variance while preventing spatial fragmentation in demarcated MZs.Geetha and ElizabehShanthi [28] have proposed methodutilised fuzzy C-Means clustering algorithm of data mining techniques used to predict crop yield in trichy district by analysing previous yield data and multilayer perception design for prediction and the results were compared to a regression algorithm for prediction.Mulyaniet al. [29] studied the productivity of rice plants in KarawangRegency;Indonesisa was classified using the Fuzzy C-Means (FCM) algorithm and Silhouette Coefficient evaluation techniques. The results of grouping for the years 2010, 2011, 2013, 2014 and 2015 were good, with only 2012 showing weak results. Zeraatpishes et al. [30] used Principal Component Analysis (PCA) and Fuzzy C-Means clustering techniques in a novel way to identify distinct soil Management Zones (MZs) that were specifically designed for commercial citrus farms in northern Iran. This integration made it possible to develop site-specific nutrient management plans, which ultimately produced economic benefits. The results showed significant variation among the six identified MZs, with MZ5 showing the best soil quality and having the lowest costs per tree.Choudharyet al. [31] used to assess the nitrogen condition of wheat crops, a link between vegetation indices obtained from an RGB camera mounted on an unmanned aerial vehicle (UAV) and actual ground data was created. The results showed that, for the two wheat types under study, there was a significant correlation between the Normalized Green-Red Difference Index (NGRDI), the Normalized Difference Vegetation Index (NDVI), and the SPAD values. Papadopoulos and Kalivas [32] have studied of two selected cotton fields in central Greece and acquired reflectance data using a commercial unmanned aerial system (UAS) before and after planting, while soil and plant properties were determined through grid sampling. Principal Component Analysis (PCA) and Fuzzy C-Means clustering algorithm were applied to catergorise sub-areas of the fields into two discrete zones per field, revealing spatial variations mostly in soil properties, which could be directly monitored through aerial reflectance data. In [33] a link between vegetation indices taken from RGB camera mounted on an Unmanned Aerial Vehicle(UAV) and real-world data was created in order to access the nitrogen status of wheat crops. The results showed a significant and remarkable link between the SPAD values, the Normalized Difference Vegetation Index(NDVI) and the Normalized Green-Red Difference Index (NGRDI) . The reliability of NGRDI as an indicator across the research domain was demonstrated by this correlation, which was clear for both varieties of wheat investigated.

3.PROPOSED WORK

The proposed model for clustering of agriculture data through Fuzzy C-Meams technique involves several steps, i.e. the data is collected from various sourcess such as sensory, satellite imagery and agricultural machinery. The data is further preprocessed to remove any missing or irrelevantinformation. Next, the Fuzzy C-Means algorithm is a popular clustering method used to group similar data points together based on the distance from mean. The fuzzy C-Means algorithm is performed over traditional clustering algorithm such as K-Means and it allows for overlapping clusters and assigns membership probabilities rather than strict classification.

After applying the Fuzzy C-Means algorithm, the resulting clusters are analysed and evaluated using various metrics such as the silhouttescore, which measures the similarity within clusters and dissimilarity between clusters. The optimal method of clusters is determined using methods such as the elbow method or silhouetteanalysis. The final step of the proposed model involves visualising the clusters using techniques such as scatter plots or heatmaps. This allows for easy interpretation and understanding of the relationship between the different agriculture data variables.

Overall, the proposed model aims to provide a reliable and accurate method for clustering agriculture data using the Fuzzy C-means technique. By doing so, it can assist the decision making process related to agriculture, such as identify crop yields, predicting weather pattrern, and determine optimal farming. The following the figure 1 represents flow of K-Means and FCM methods.



(a) K-Means Method **(b)** FCM Method

Figure 1. Flow of Clustering Methods

BanshidharChoudhary/Afr.J.Bio.Sc. 6(13)(2024).3312-3329

4.RESULTS AND DISCUSSION

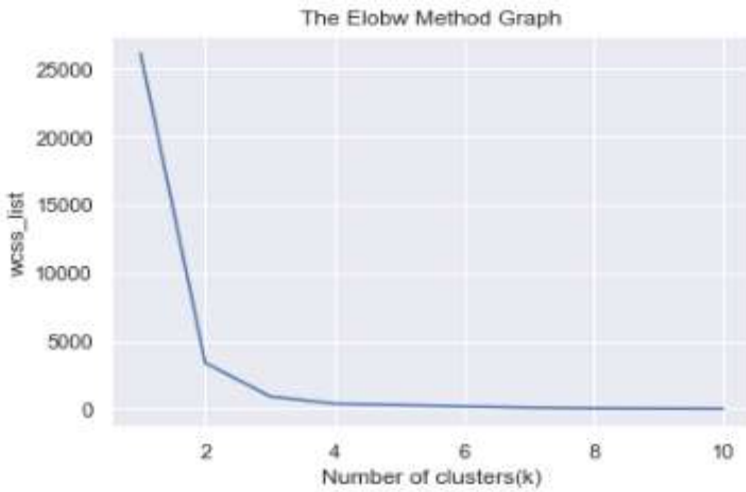
On the basis on proposed models, let us discuss some interesting results . The table1 displays information on crop production in India from 2010-11 to 2020-21. The crop are represented by the respective attribute values, where 1 represents Rice, 2 represents Wheat, 3 represents Coarse, 4 represents Jowyar , 5 represents Bajra, 6 represents Maize, 7 represents total Pulse, 8 represents Gram, 9 represents Arhare,10 represents Masure, 11 represents oil seeds, represents Grounds nut, 12 represents Grounds nut, 13 represents Mustered and 14 represents Soybean production

Table 1. Data of Crop Production

Crop	2010-11	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21
1	95	105	105	106	105	104	109	112	116	118	122
2	86	94	93	95	86	92	98	99	103	107	109
3	43	42	40	43	42	38	43	46	43	47	51
4	7	5	5	5	5	4	4	4	3	4	4
5	10	10	8	9	9	8	9	9	8	10	10
6	21	21	22	24	24	22	25	28	25	28	31
7	18	17	18	19	17	16	23	25	22	23	25
8	8	7	8	9	7	7	9	11	9	11	11
9	2	2	3	3	2	2	4	4	3	3	4
10	1	1	1	1	1	1	1	1	1	1	1
11	32	29	30	32	27	25	31	31	31	32	36
12	8	6	4	9	7	6	7	9	6	9	10
13	8	6	8	7	6	6	7	8	9	9	10
14	12	12	14	11	10	8	13	10	13	11	12

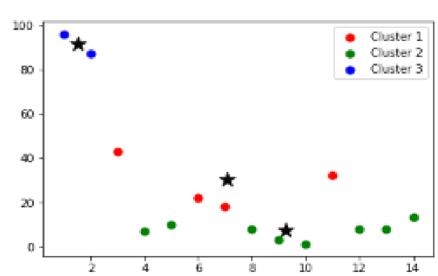
By applying the Fuzzy C-Means algorithm using Python, the crop production data in the table1 was analyzed with a value of k=3. The elbow method was used to determine the optimal number of clusters and the results are shown in the figure 2(a). The FCM clustering of crop data is shown in figure 2(b) , where cluster#1 includes RiceandWheat crops, cluster#2 includes Total Pulse, Oil seed, Jowar, Bajara, Grame, Arhare, Masure, Ground Nut, Mustard Oil and Soybeans. When k=4, the fuzzy cluster was represented in figure 3(a), which shows that only crop 3(Coarse) belongs toits own cluster.

The use of Fuzzy C-Means algorithm on crop production data can be provided insight for decision making related to agriculture, such as identifying the best crop for specific regions or predicting crop yields.



(a). Elbow Method

	Crop	2010-11	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21
0	1	96	105	105	107	105	104	110	113	116	119	122
1	2	87	95	94	96	87	92	99	100	104	108	109
2	3	43	42	40	43	43	39	44	47	43	48	51
3	4	7	6	5	6	5	4	5	5	3	5	5
4	5	10	10	9	9	9	8	10	9	9	10	11
5	6	22	22	22	24	24	23	26	29	26	29	32
6	7	18	17	18	19	17	16	23	25	22	23	26
7	8	8	8	9	10	7	7	9	11	10	11	12
8	9	3	3	3	3	3	3	5	4	3	4	4
9	10	1	1	1	1	1	1	1	2	1	1	1
10	11	32	30	31	33	28	25	31	31	32	32	36
11	12	8	7	5	10	7	7	7	9	7	10	10
12	13	8	7	8	8	6	7	8	8	9	9	10
13	14	13	12	15	12	10	9	13	11	13	11	13



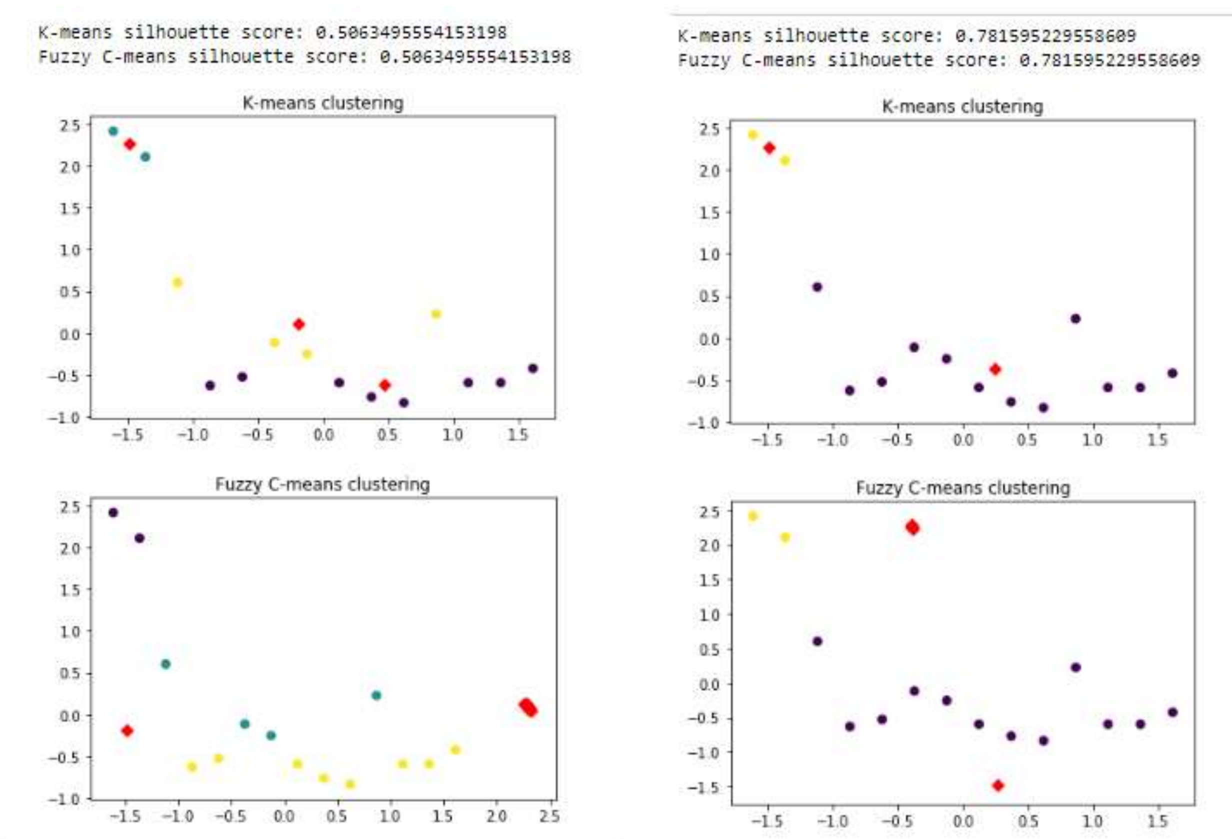
(b). FCM Clustering

Figure 2. Representation of Clustering Method

Performance comparison of K-Means and FCM method were carried out by varying the value of k as 2,3 and 4 and calculating the Silhouette Score for each case. The results are presented in the table 2, and it was observed that when k=2 or 3, both K-Means and FCM have the same Silhouette Score. However, when k=4, K-Means has a Silhouette Score of 0.3945, while FCM has a slightly higher.

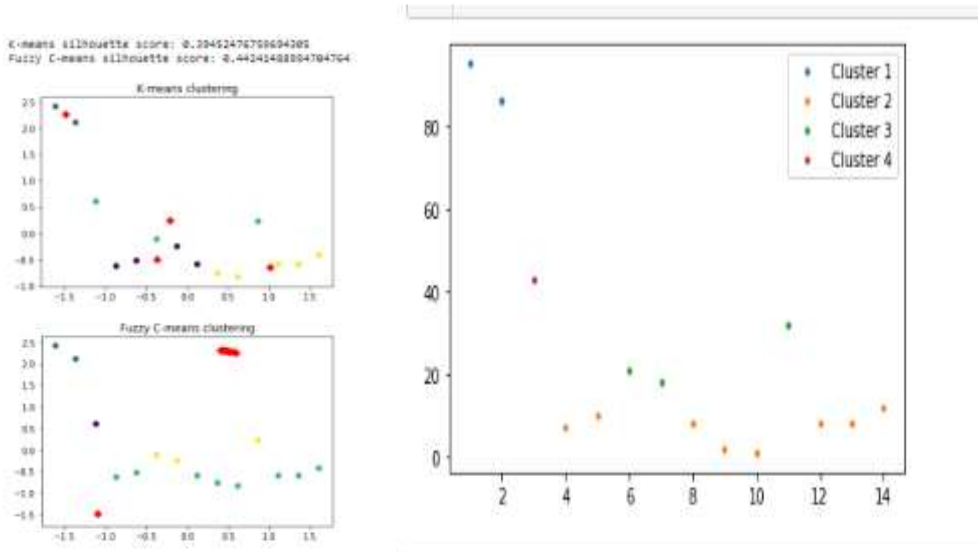
Table 2. Performance of K-Means and FCM Methods

Value of K	K-Means(Silhouette Score)	FCM(Silhouette Score)
2	0.781595229558609	0.781595229558609
3	0.5063495554153198	0.5063495554153198
4	0.39452476759694305	**0.44241488994704764**



(a). Comparison at K=2 **(b).**Comparison at K=3.

Figure 3. Comparison of K-Means and Fuzzy C-Means Clustering



(a). Comparison at K=4 **(b).**Fuzzy Cluster of Crop Data

Figure 4.Comparison of Clustering Algorithms

Using the R language, we performed Fuzzy C-Means clustering on the crop production data in table 1. The resulting data is displayed in the figure 5(a) known as Head(df) and figure 5(b) known as cluster center of crop production, which displays the cluster centers for k=3. From figure 5(b), we can concluded that cluster#1 contains crops 1,2 and 3 (i.e. Rice, Wheat and Coarse), cluster#2 contains crop 6,7 and 11 (i.e. Maize, Total Pulse and Oil Seed) and cluster#3 contains crop 4,5,8,9,10,12,13 and 14 (i.e. Jowar, Bajara, Cereals, Arhar, Masure, Ground Nut, Mustard and Soybeans).

```
> setwd("F:/lab")
> df=read.csv("bddatafile.csv")
> str(df)
'data.frame':   14 obs. of  12 variables:
 $ Crop     : int  1 2 3 4 5 6 7 8 9 10 ...
 $ X2010.11: num  96 86.9 43.4 7 10.4 ...
 $ X2011.12: num  105.3 94.88 42.01 5.98 10.28 ...
 $ X2012.13: num  105.23 93.51 40.04 5.28 8.74 ...
 $ X2013.14: num  106.65 95.85 43.29 5.54 9.25 ...
 $ X2014.15: num  105.48 86.53 42.86 5.45 9.18 ...
 $ X2015.16: num  104.41 92.29 38.52 4.24 8.07 ...
 $ X2016.17: num  109.7 98.51 43.77 4.57 9.73 ...
 $ X2017.18: num  112.76 99.87 46.97 4.8 9.21 ...
 $ X2018.19: num  116.48 103.6 43.06 3.48 8.66 ...
 $ X2019.20: num  118.87 107.86 47.75 4.77 10.36 ...
 $ X2020.21: Factor w/ 14 levels "1.45","10.11",..: 8 5 14 13 4 10 9 6 12 1 ...
> head(df)
  Crop X2010.11 X2011.12 X2012.13 X2013.14 X2014.15 X2015.16 X2016.17 X2017.18 X2018.19
1    1    95.98   105.30   105.23   106.65   105.48   104.41   109.70   112.76   116.48
2    2    86.87    94.88    93.51    95.85    86.53    92.29    98.51    99.87   103.60
3    3    43.40    42.01    40.04    43.29    42.86    38.52    43.77    46.97    43.06
4    4     7.00     5.98     5.28     5.54     5.45     4.24     4.57     4.80     3.48
5    5    10.37    10.28     8.74     9.25     9.18     8.07     9.73     9.21     8.66
6    6    21.73    21.76    22.26    24.26    24.17    22.57    25.90    28.75    25.72
  X2019.20 X2020.21
1   118.87   122.27
2   107.86   109,52
3    47.75    51.45
4     4.77     4.78
5    10.36    10.86
6    28.77    31.51
```

(a). Head (df)

BanshidharChoudhary/Afr.J.Bio.Sc. 6(13)(2024).3312-3329

```
+ }
Cluster 1 :Crop: 1 2 3X2010.11: 95.98 86.87 43.4X2011.12: 105.3 94.88 42.01X2012.13: 105.23 93.51 40.04X2013.14:
 106.65 95.85 43.29X2014.15: 105.48 86.53 42.86X2015.16: 104.41 92.29 38.52X2016.17: 109.7 98.51 43.77X2017.18:
 112.76 99.87 46.97X2018.19: 116.48 103.6 43.06X2019.20: 118.87 107.86 47.75X2020.21: 8 5 14

Cluster 2 :Crop: 6 7 11X2010.11: 21.73 18.24 32.48X2011.12: 21.76 17.09 29.8X2012.13: 22.26 18.34 30.94X2013.14:
 24.26 19.25 32.75X2014.15: 24.17 17.15 27.51X2015.16: 22.57 16.32 25.25X2016.17: 25.9 23.13 31.28X2017.18: 28.7
5 25.42 31.46X2018.19: 25.72 22.08 31.52X2019.20: 28.77 23.03 32.22X2020.21: 10 9 11

Cluster 3 :Crop: 4 5 8 9 10 12 13 14X2010.11: 7 10.37 8.22 2.86 0.94 8.26 8.18 12.74X2011.12: 5.98 10.28 7.7 2.6
5 1.06 6.96 6.6 12.21X2012.13: 5.28 8.74 8.83 3.02 1.13 4.7 8.03 14.67X2013.14: 5.54 9.25 9.75 3.17 1.02 9.71 7.
88 11.86X2014.15: 5.45 9.18 7.33 2.81 1.04 7.4 6.25 10.37X2015.16: 4.24 8.07 7.06 2.56 0.98 6.73 6.8 8.57X2016.1
7: 4.57 9.73 9.38 4.87 1.22 7.46 7.92 13.16X2017.18: 4.8 9.21 11.38 4.29 1.62 9.25 8.43 10.93X2018.19: 3.48 8.66
 9.94 3.32 1.23 6.73 9.26 13.27X2019.20: 4.77 10.36 11.08 3.89 1.1 9.95 9.12 11.23X2020.21: 13 4 6 12 1 3 2 7

> |
```

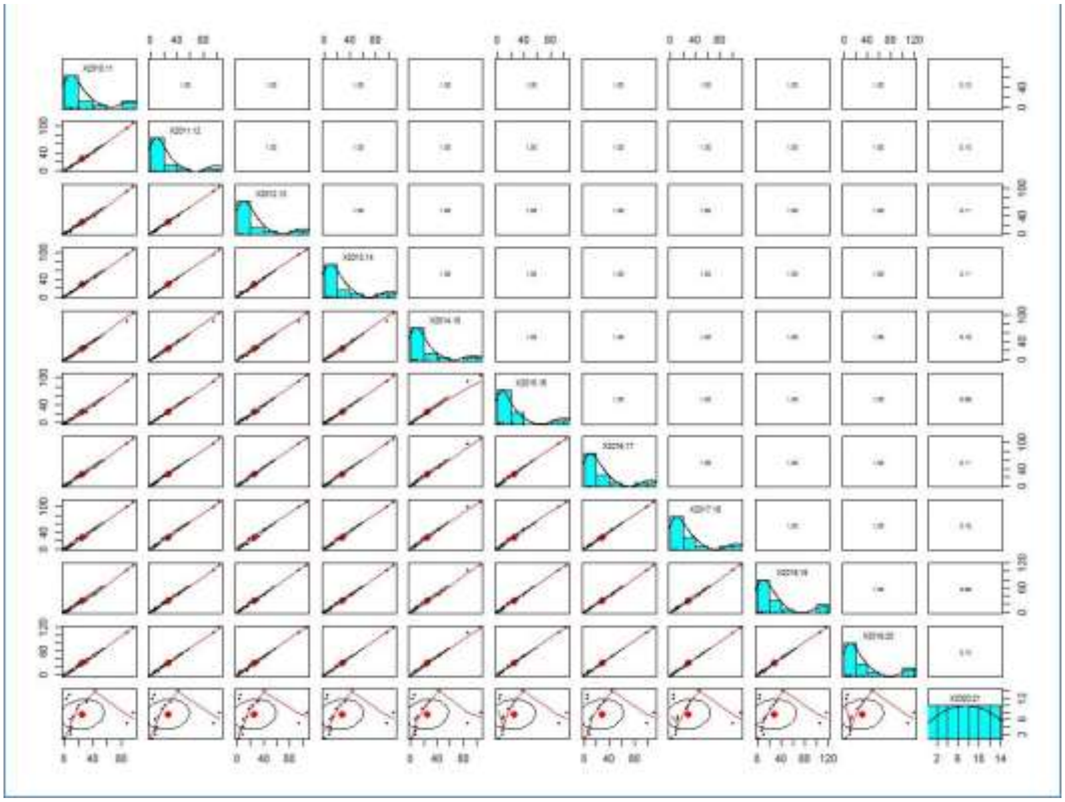
(b). Cluster Centre

Figure 5. Resulting Data for Head and Cluster Centre

To perform the FCM clustering , we paired the data set in table1, which generated the pairing information displayed in figure 6(a) known as pairing data and the corresponding pairing graph

```
> pairs(df, col=df[,5])
> cor(df[,1:12])
                Crop   X2010.11   X2011.12   X2012.13   X2013.14   X2014.15   X2015.16
Crop       1.0000000 -0.6694620 -0.6733819 -0.6586697 -0.6679202 -0.6831197 -0.6796846
X2010.11 -0.6694620  1.0000000  0.9983861  0.9972906  0.9986762  0.9963711  0.9956514
X2011.12 -0.6733819  0.9983861  1.0000000  0.9991491  0.9991894  0.9978946  0.9987822
X2012.13 -0.6586697  0.9972906  0.9991491  1.0000000  0.9984832  0.9970943  0.9981190
X2013.14 -0.6679202  0.9986762  0.9991894  0.9984832  1.0000000  0.9979849  0.9986008
X2014.15 -0.6831197  0.9963711  0.9978946  0.9970943  0.9979849  1.0000000  0.9980111
X2015.16 -0.6796846  0.9956514  0.9987822  0.9981190  0.9986008  0.9980111  1.0000000
X2016.17 -0.6673671  0.9971775  0.9987038  0.9988581  0.9991080  0.9972605  0.9983121
X2017.18 -0.6794699  0.9960809  0.9969983  0.9966570  0.9985140  0.9974634  0.9977536
X2018.19 -0.6583334  0.9962463  0.9985823  0.9992772  0.9986900  0.9971952  0.9988288
X2019.20 -0.6737863  0.9968781  0.9984208  0.9979460  0.9992514  0.9974690  0.9990486
X2020.21 -0.6696413  0.9974806  0.9977547  0.9974592  0.9992394  0.9976999  0.9978564
            X2016.17   X2017.18   X2018.19   X2019.20   X2020.21
Crop      -0.6673671 -0.6794699 -0.6583334 -0.6737863 -0.6696413
X2010.11   0.9971775  0.9960809  0.9962463  0.9968781  0.9974806
X2011.12   0.9987038  0.9969983  0.9985823  0.9984208  0.9977547
X2012.13   0.9988581  0.9966570  0.9992772  0.9979460  0.9974592
X2013.14   0.9991080  0.9985140  0.9986900  0.9992514  0.9992394
X2014.15   0.9972605  0.9974634  0.9971952  0.9974690  0.9976999
X2015.16   0.9983121  0.9977536  0.9988288  0.9990486  0.9978564
X2016.17   1.0000000  0.9990504  0.9995482  0.9993305  0.9991346
X2017.18   0.9990504  1.0000000  0.9983944  0.9994062  0.9996082
X2018.19   0.9995482  0.9983944  1.0000000  0.9992056  0.9986069
X2019.20   0.9993305  0.9994062  0.9992056  1.0000000  0.9995247
X2020.21   0.9991346  0.9996082  0.9986069  0.9995247  1.0000000
```

(a). Pairing Data



(b). Pairing Graph

Figure 6. Representataion of Pairing

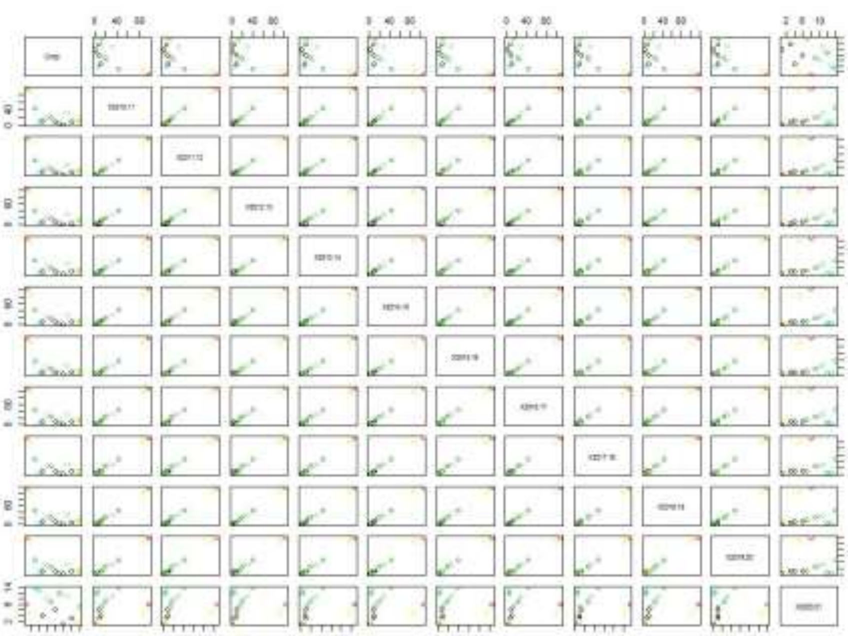
displayed in figure 6(b)Know as pairing graph.. The pairing graph shows the degree of similarity between each pair of data points, which is used to cluster the data into group.

In R language , we can copmpare the K-Means and Fuzzy C-Means clustering algorithms. To do so we use 'fanny' function to perform K-Means clustering on the crop production data of table 1. The results are displayed in 7(a) as membership of each data point. We can calculate within-cluster sum of squares for the K-Means clustering results and save the fuzzy C-Means clustering in the 'fc-cluster' object using the 'fanny( )' function. The membership element along with thefuzziness coefficients are deployed with closest hard clustering.

```
> print(km_cluster$tot.withinss)
[1] 5305.625
>
> # Plot the k-means clustering result
> plot(crop_data[, -1], col = km_cluster$cluster, main = "K-means Clustering")
>
> # Run fuzzy c-means clustering
> fc_cluster <- fanny(crop_data[, -1], k, metric = "euclidean", stand = FALSE)
>
> # Print the fuzzy c-means clustering result
> print(fc_cluster)
Fuzzy Clustering object of class 'fanny' :
m.ship.expon.        2
objective     111.6886
tolerance        1e-15
iterations          21
converged            1
maxit              500
n                   14
Membership coefficients (in %, rounded):
      [,1] [,2] [,3]
 [1,]   93    4    3
 [2,]   92    5    3
 [3,]   13   67   21
 [4,]    2    8   90
 [5,]    1    8   90
 [6,]    2   86   12
 [7,]    4   69   27
 [8,]    1    8   91
 [9,]    2   10   87
[10,]    3   14   82
[11,]    3   87   10
[12,]    1    6   93
[13,]    1    6   93
[14,]    3   20   78
Fuzzyness coefficients:
dunn_coeff normalized
 0.7592653  0.6388980
Closest hard clustering:
 [1] 1 1 2 3 3 2 2 3 3 3 2 3 3 3

Available components:
 [1] "membership"  "coeff"       "memb.exp"    "clustering"  "k.crisp"     "objective"
```

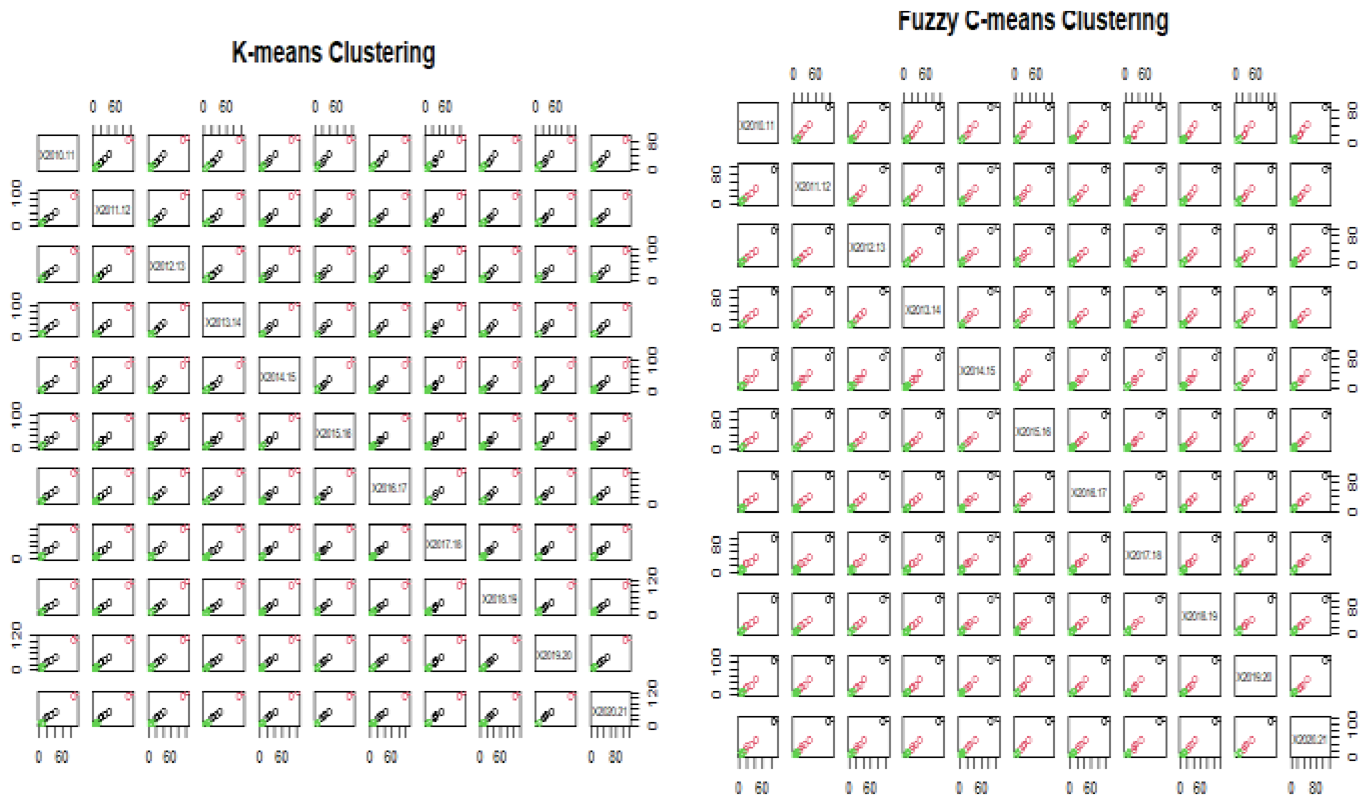
(a) Crop Data



(b) Production Plot

Figure 7. Representation of Membership and Production Plot

We can then plot the crop production data in the figure 7(b), which shows the K-Means cluster in figure 8(a) and the Fuzzy C-Means cluster in figure 8(b).



(a). K-Means Cluster

(b). Fuzzy C-Means.

Figure 8. Clustering of Crop Data

Here, we are considering another agriculture data set,agri2.csv, which is shown in the figure 9(b). Figure 9(a) is the same as the table 1 data. For the consideration of K-Means and FCM of two agriculture datasets, the results are given in the figure 9.

```
> setwd("F:/lab")
> df=read.csv("bddatafile.csv")
> str(df)
'data.frame':	14 obs. of  12 variables:
 $ Crop    : int  1 2 3 4 5 6 7 8 9 10 ...
 $ X2010.11: num  96 86.9 43.4 7 10.4 ...
 $ X2011.12: num  105.3 94.88 42.01 5.98 10.28 ...
 $ X2012.13: num  105.23 93.51 40.04 5.28 8.74 ...
 $ X2013.14: num  106.65 95.85 43.29 5.54 9.25 ...
 $ X2014.15: num  105.48 86.53 42.86 5.45 9.18 ...
 $ X2015.16: num  104.41 92.29 38.52 4.24 8.07 ...
 $ X2016.17: num  109.7 98.51 43.77 4.57 9.73 ...
 $ X2017.18: num  112.76 99.87 46.97 4.8 9.21 ...
 $ X2018.19: num  116.48 103.6 43.06 3.48 8.66 ...
 $ X2019.20: num  118.87 107.86 47.75 4.77 10.36 ...
 $ X2020.21: Factor w/ 14 levels "1.45","10.11",..: 8 5 14 13 4 10 9 6 12 1 ...
> head(df)
  Crop X2010.11 X2011.12 X2012.13 X2013.14 X2014.15 X2015.16 X2016.17 X2017.18 X2018.19
1    1    95.98   105.30   105.23   106.65   105.48   104.41   109.70   112.76   116.48
2    2    86.87    94.88    93.51    95.85    86.53    92.29    98.51    99.87   103.60
3    3    43.40    42.01    40.04    43.29    42.86    38.52    43.77    46.97    43.06
4    4     7.00     5.98     5.28     5.54     5.45     4.24     4.57     4.80     3.48
5    5    10.37    10.28     8.74     9.25     9.18     8.07     9.73     9.21     8.66
6    6    21.73    21.76    22.26    24.26    24.17    22.57    25.90    28.75    25.72
  X2019.20 X2020.21
1   118.87   122.27
2   107.86   109,52
3    47.75    51.45
4     4.77     4.78
5    10.36    10.86
6    28.77    31.51
```

(a) as bddatafile.csv

```
'data.frame':	49 obs. of  7 variables:
 $ Crop                 : int  0 0 0 0 0 1 1 1 1 1 ...
 $ State                : int  1 2 3 4 5 5 6 4 3 7 ...
 $ Cost.of.Cultivation  : num  9794 10593 13469 17052 17131 ...
 $ Cost.of.Cultivation.1: num  23077 16529 19552 24172 25270 ...
 $ Cost.of.Production   : num  1942 2172 1898 3671 2776 ...
 $ Yield                : num  9.83 7.47 9.59 6.42 8.72 ...
 $ X                    : logi  NA NA NA NA NA NA ...
> data
   Crop State Cost.of.Cultivation Cost.of.Cultivation.1 Cost.of.Production   Yield
1     0     1             9794.05              23076.74            1941.55    9.83
2     0     2            10593.15              16528.68            2172.46    7.47
3     0     3            13468.82              19551.90            1898.30    9.59
4     0     4            17051.66              24171.65            3670.54    6.42
5     0     5            17130.55              25270.26            2775.80    8.72
6     1     5            23711.44              33116.82            2539.47   12.69
7     1     6            29047.10              50828.83            2003.76   24.39
8     1     4            29140.77              44756.72            2509.99   17.83
9     1     3            29616.09              42070.44            2179.26   19.05
10    1     7            29918.97              44018.18            2127.35   19.90
11    2     8             8552.69              12610.85            1691.66    6.83
12    2     9             9803.89              16873.17            1551.94   10.29
13    2     1            12833.04              21618.43            1882.68   10.93
14    2     5            12985.95              18679.33            2277.68    8.05
15    2     4            14421.98              26762.09            1559.04   16.69
16    3     2            13647.10              17314.20            3484.01    4.71
17    3     4            21229.01              30434.61            2554.91   11.97
18    3    10            22507.86              30393.66            2358.00   11.98
19    3     3            22951.28              30114.45            1918.92   13.45
20    3     5            26078.66              32683.46            3207.35    9.33
21    4    11            13513.92              19857.70             404.43   42.95
22    4     2            13792.85              20671.54             581.69   31.10
23    4     8            14421.46              19810.29             658.77   23.56
24    4     1            15635.43              21045.11            1387.36   13.70
25    4     4            25687.09              37801.85             840.58   42.68
26    5    12             5483.54               8266.98            2614.14    3.01
27    5     8             6204.23               9165.59            2068.67    4.05
28    5     2             6440.64               7868.64            5777.48    1.32
29    5     4             6684.18              13209.32            2228.97    5.90
30    5     5            10780.76              15371.45            2261.24    6.70
31    6     1            17022.00              28144.50             732.62   36.61
32    6    12            17478.05              25909.05             715.04   32.42
33    6    13            24731.06              33046.12             731.25   39.04
34    6     6            25154.75              45291.24             669.86   67.41
35    6     4            29664.84              48450.20             789.90   56.00
36    7     9             8686.43              17705.93            1279.60   12.94
37    7     8            11385.70              19259.84            1341.29   13.54
38    7     1            12774.41              22560.30            1595.56   13.57
39    7     3            13740.64              19083.55            1610.40   11.61
40    7     7            14715.27              27507.54            1251.12   19.94
41    8     1            24538.32              45239.51              93.64  448.89
42    8     2            55655.44              86765.77              86.53  986.21
```

(b) as agri2.csv

Figure 9. Representation of Data for K-Means and FCM

BanshidharChoudhary/Afr.J.Bio.Sc. 6(13)(2024).3312-3329

In table 3, we are providing the silhouette Score of K-Means and FCM by using both Python and R language at different value of K. We concluded that the optimized FCM value is at the value of k where the Silhouette Score is the highest.

Table 3. K-Means and FCM Silhouette Score for agri2.csv and bddatafile.csv Dataset

S.No.	Data Set	Language	C(No of Cluster)	m(Fuzziness) /PCA	K-Means Silhouette Score	Fuzzy C-Mean Silhouette Score	Conclusion
1	AGRI2.CSV	R	2	1.5	.628231	**.7936738**	FCM optimize at **K=2** and **K= 4**
2			3		.6170364	.6170364	
3			4		.5061651	**.5831066**	
4			5		.5818557	.5394483	
5			6		.5461024	.503695	
6	AGRI2.CSV	Python	2	PCA	.784	**.784**	FCM optimize at **K=2**
7			3		.456	.447	
8			4		.474	.455	
9			5		.509	.495	
10			6		.511	.510	
11	BDDATAFILE1 .CSV	R	2	1.5	.8277224	.8277224	FCM optimize at **K=5** and better than K-Means
12			3		.6446251	.6446251	
13			4		.586626	.586626	
14			5		.4403872	**.5311954**	
15			6		.4877484	.4877484	
16	BDDATAFILE1 .CSV	Python	2	PCA	.829	.829	FCM optimize at **K=6**
17			3		.695	.695	
18			4		.653	.653	
19			5		.541	.541	
20			6		.437	**.494**	

5. CONCLUSIONS

From the above , it is concluded that the clustering of agriculture data through the Fuzzy C-Means technique can provides various insight into agriculture production and helps make to data – driven decision. By implementing the proposed model, farmers and policy maker can give a better understsnding of various factors and policy makers can gain a better uunderstanding of the various factors that affect crop yield such as climate, soil quality and cost of production. Additionaly, this approach can help ,to identify the patterns and trends in the data that may not be immediately apperent and guide the development of more targated intervention and policies to improv, agriculture outcomes.Overall, the Fuzzy C-Means technique is a powerful technique for clustering agricluture data and has the potential to revolutionalise the agriculture producers and management.

REFERENCES

[1] Córdoba, M., Bruno, C., Costa, J., &Balzarini, M. (2013). Subfield management class delineation using cluster analysis from spatial principal components of soil variables. *Computers and Electronics in Agriculture*, 97. https://doi.org/10.1016/j.compag.2013.05.009

[2] Bansod, B. S., & Pandey, O. P. (2013). An application of PCA and fuzzy C-means to delineate management zones and variability analysis of soil. *Eurasian Soil Science*, 46(5). https://doi.org/10.1134/S1064229313050165

[3] Speranza, E. A., Ciferri, R. R., Grego, C. R., & Vicente, L. E. (2014). A Cluster-Based Approach to Support the Delineation of Management Zones in Precision Agriculture. In 2014 *IEEE 10th International Conference on e-Science* (Vol. 1, pp. 119–126). IEEE. https://doi.org/10.1109/eScience.2014.42

[4] Chang, D., Zhang, J., Zhu, L., Ge, S. H., Li, P. Y., & Liu, G. S. (2014). Delineation of management zones using an active canopy sensor for a tobacco field. *Computers and Electronics in Agriculture*, 109. https://doi.org/10.1016/j.compag.2014.09.019

[5] Patel, H., & Patel, D. (2014). A Brief survey of Data Mining Techniques Applied to Agricultural Data. *International Journal of Computer Applications*, 95(9). https://doi.org/10.5120/16620-6472

[6] Singh, S., Ambegaokar, S., Champawat, K. S., Gupta, A., & Sharma, S. (2015). Time series analysis of clustering high dimensional data in precision agriculture. In 2015 *International Conference on Innovations in Information, Embedded and Communication Systems* (ICIIECS) (pp. 1–8). IEEE. https://doi.org/10.1109/ICIIECS.2015.7193192

[7] Bazzi, C. L., de Souza, E. G., Konopatzki, M. R., Pereira Nóbrega, L. H., & Uribe-Opazo, M. A. (2015). Management zones applied to pear orchard. *Journal of Food, Agriculture and Environment*, 13(1).

[8] Rodrigues, M. S., & Corá, J. E. (2015). Management zones using fuzzy clustering based on spatial-temporal variability of soil and corn yield. *Engenharia Agricola*, 35(3). https://doi.org/10.1590/1809-4430-Eng.Agric.v35n3p470-483/2015

[9] Marcos, S. R., Matheus, C. S., Amelia, L. P. U. ocirc a, Alex, X. da S. M. de M., Italo, H. L. C., & Augusto, M. N. L. (2015). Delineation of management zones based on soil physical attributes in an irrigated guava field in the Semi-Arid region, Brazil. *African Journal of Agricultural Research*, 10(45). https://doi.org/10.5897/ajar2015.9901

[10] Speranza, E. A., Ciferri, R. R., &Ciferri, C. D. A. (2016). Clustering approaches and ensembles applied in the delineation of management classes in precision agriculture. In *Proceedings of the Brazilian Symposium on GeoInformatics*(Vol. 2016-November).

[11] Gavioli, A., de Souza, E. G., Bazzi, C. L., Guedes, L. P. C., &Schenatto, K. (2016). Optimization of management zone delineation by using spatial principal components. *Computers and Electronics in Agriculture*, 127. https://doi.org/10.1016/j.compag.2016.06.029

[12] Ghosh, S., Nandy, S., Patra, S., Kushwaha, S. P. S., Senthil Kumar, A., &Dadhwal, V. K. (2017). Land Cover Classification Using ICESat/GLAS Full Waveform Data. *Journal of the Indian Society of Remote Sensing*, 45(2). https://doi.org/10.1007/s12524-016-0602-5

[13] Tamiminia, H., Homayouni, S., McNairn, H., & Safari, A. (2017). A particle swarm optimized kernel-based clustering method for crop mapping from multi-temporal polarimetric L-band SAR observations. *International Journal of Applied Earth Observation and Geoinformation*, 58. https://doi.org/10.1016/j.jag.2017.02.010

[14] Sindhu, S., & Sindhu, D. (2017). Role of Data Mining Techniques in Agriculture Improvement. *IJARCCE*, 6(5). https://doi.org/10.17148/ijarcce.2017.65126

[15] Abdullahi1, A., Muhammad, M. T., Suleiman, J., & Sokoto1, R. M. (2018). Isolation and Identification of Bacteria Associated with Aerial Part of Rice Plant from Kware Lake. *Asian Journal of Research in Botany*, 1(2).

[16] mohajer, reza, &Salehi, M. H. (2018). American Soil Taxonomy Compared to World Reference Base for Expressing Environmental Pollution, a Case Study: *Lenjanat Region of Isfahan. Water and Soil*, 31(6).

[17] Camicia, R. G. da M., Maggi, M. F., de Souza, E. G., Bazzi, C. L., Konopatzki, E. A., Michelon, G. K., &Pinheiro, J. B. S. (2018). Productivity of soybean in management zones with application of different sowing densities. *Ciencia Rural*, 48(12). https://doi.org/10.1590/0103-8478cr20180532

[18] Heupel, K., Spengler, D., &Itzerott, S. (2018). A Progressive Crop-Type Classification Using Multitemporal Remote Sensing Data and Phenological Information. PFG - *Journal of Photogrammetry, Remote Sensing and Geoinformation Science,* 86(2). https://doi.org/10.1007/s41064-018-0050-7

[19] Cillis, D., Pezzuolo, A., Marinello, F., &Sartori, L. (2018). Field-scale electrical resistivity profiling mapping for delineating soil condition in a nitrate vulnerable zone. *Applied Soil Ecology*, 123. https://doi.org/10.1016/j.apsoil.2017.06.025

[20] Rajput, A., &Kumaravelu, V. B. (2019). Scalable and sustainable wireless sensor networks for agricultural application of Internet of things using fuzzy c-means algorithm. Sustainable Computing: *Informatics and Systems*, 22. https://doi.org/10.1016/j.suscom.2019.02.003

[21] Zhao, S., Chen, G., Fu, S., & Xiao, E. (2019). Study on Precision Fertilization Model Based on Fusion Algorithm of Cluster and RBF Neural Network. In IFIP *Advances in Information and Communication Technology* (Vol. 546, pp. 56–66). https://doi.org/10.1007/978-3-030-06179-1_6

[22] Yassine, S., Najib, E. K., & Fatima, L. (2019). Dynamic Cluster Head Selection Method for Wireless Sensor Network for Agricultural Application of Internet of Things based Fuzzy C-means Clustering Algorithm. In 2019 *7th Mediterranean Congress of Telecommunications (CMT)* (pp. 1–9). IEEE. https://doi.org/10.1109/CMT.2019.8931313

[23] Pavan Kumar, A. V. S., &Bhramaramba, R. (2019). Optimized multi-attribute water quality resource assessment in agriculture sector. *International Journal of Scientific and Technology Research*, 8(11).

[24] Brubeck-Hernandez, F., Vladimirova, T., Pooley, M., Thompson, R., & Knight, B. (2019). Zone Management in Precision Agriculture Using Satellite Imagery. In 2019 *NASA/ESA Conference on Adaptive Hardware and Systems (AHS)* (pp. 65–71). IEEE. https://doi.org/10.1109/AHS.2019.00006

[25] Liu, B., He, S., He, D., Zhang, Y., &Guizani, M. (2019). A Spark-Based Parallel Fuzzy c -Means Segmentation Algorithm for Agricultural Image Big Data. *IEEE Access*, 7. https://doi.org/10.1109/ACCESS.2019.2907573

[26] Bai, C., Shi, B., Liu, F., &Sarkis, J. (2019). Banking credit worthiness: Evaluating the complex relationships. *Omega (United Kingdom)*, 83. https://doi.org/10.1016/j.omega.2018.02.001

[27] Gavioli, A., de Souza, E. G., Bazzi, C. L., Schenatto, K., &Betzek, N. M. (2019). Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. *Biosystems Engineering*, 181. https://doi.org/10.1016/j.biosystemseng.2019.02.019

[28] Geetha, M. C. S., & Elizabeth Shanthi, I. (2020). Forecasting the crop yield production in trichy district using fuzzy C-Means Algorithm and Multilayer Perceptron (MLP). *International Journal of Knowledge and Systems Science,* 11(3). https://doi.org/10.4018/IJKSS.2020070105

[29] Mulyani, S., Sari, B. N., &Ridha, A. A. (2020). Clustering Productivity of Rice in Karawang Regency Using the Fuzzy C-Means Method. *Indonesian Journal of Artificial Intelligence and Data Mining*, 3(2).
[30] Zeraatpisheh, M., Bakhshandeh, E., Emadi, M., Li, T., & Xu, M. (2020). Integration of PCA and Fuzzy Clustering for Delineation of Soil Management Zones and Cost-Efficiency Analysis in a Citrus Plantation. *Sustainability*, 12(14), 5809. https://doi.org/10.3390/su12145809
[31] Choudhary, S. S., Biswal, S., Saha, R., & Chatterjee, C. (2021). A non-destructive approach for assessment of nitrogen status of wheat crop using unmanned aerial vehicle equipped with RGB camera. *Arabian Journal of Geosciences*, 14(17). https://doi.org/10.1007/s12517-021-08139-3
[32] Papadopoulos, A. V., &Kalivas, D. P. (2021). Assessing soil and crop characteristics at sub-field level using unmanned aerial system and geospatial analysis. *Sustainability (Switzerland)*, 13(5). https://doi.org/10.3390/su13052855
[33] Huang, C., & Chen, Y. (2021). Agricultural business and product marketing effected by using big data analysis in smart agriculture. *ActaAgriculturaeScandinavica Section B: Soil and Plant Science*, 71(9). https://doi.org/10.1080/09064710.2021.1967439