FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 5, May, 2025, pp 136 - 143
DOI: https://doi.org/10.33003/fjs-2025-0905-3658

APPLICATION OF K-MEANS AND K-MEDOID CLUSTERING TECHNIQUES ON RICE YIELDS IN NIGERIA

***Ismail, N. Y., Abdullahi, U. K., Musa, T., Garba, J. and Abdulkarim Hafsat**

Department of Statistics, Ahmadu Bello University Zaria. Kaduna State, Nigeria

*Corresponding authors' email: nyismail01@gmail.com

ABSTRACT

This paper explores the application of two clustering algorithms, K-means and K-medoid to classify rice yields across 36 Nigerian states and the Federal Capital Territory (FCT). The study aims to identify natural groupings within the data to classify the states base on their similarity in the rice yields. The findings reveal that the K-means and K-medoid clustering effectively grouped rice and maize yields into six clusters each. The silhouette width indicated that K-medoid out performed K-means in cluster quality with an average silhouette width of 0.42 and for rice compared to K-means with silhouette width of 0.39 and. The analysis highlighted significant clusters, such as clusters three and five for rice using K-means and K-medoid which represent regions with similar average crop yields. These insights underline the importance of targeted governmental interventions to improve agricultural productivity by focusing on areas with average yields. The results also suggest that strategic investments in infrastructure, agricultural inputs, and policy reforms are crucial for boosting productivity and reducing poverty. Overall, the paper concludes that K-medoid is the superior clustering technique, delivering the highest silhouette width and the most accurate classifications for both crops, with the fewest misclassifications. This research provides a valuable framework for regional agricultural planning and resource allocation in Nigeria. The paper recommends that the government should pay attention on allocating the scarce resources to the consistency clusters along with policy review in favor of smallholder farmers through access and timely for all important farm inputs in future.

Keywords: Multivariate techniques, Clustering algorithms, K-means, K-medoid, Sil-width

INTRODUCTION

Cluster Analysis (CA) is a multivariate technique used to organize a set of multivariate data (observations or objects) into groups known as clusters. Observations within each cluster are similar to one another, while observations in different clusters are dissimilar. It is an exploratory method that aids in understanding the underlying patterns within a dataset (Wierzchoń & Kłopotek, 2018). The fundamental idea behind cluster analysis is to group cases that are similar based on their characteristics across recorded variables. Through this method, dense and sparse regions within the data distribution can be identified, allowing for the recognition of different distribution patterns. The concept of similarity and dissimilarity is central to cluster analysis. Various distance or similarity measures can be used to evaluate relationships between data points. In this study, the Euclidean distance measure is employed. Cluster analysis has been widely applied in numerous research fields, including criminology, sociology, law, management, counseling, psychology, and statistical literatures.

K-means clustering algorithms developed by Macqueen (1967) is one of the most widely used algorithm for splitting dataset into number of k clusters, in which k denotes the number of clusters. It partitions items or objects in multiple clusters, such that items in the same cluster are similar to each other with high intra class similarity; while items from different clusters are distinct with low inter class similarity. In k-means algorithm, every cluster is denoted by its centroid which is defined as the mean of points within the given cluster. In k-means clustering algorithm, the first step is to determine the number of clusters which will be obtained as the final result, which is the parameter *k*. Then k items or objects are randomly selected as centroids of the cluster. All remaining items (objects) are assigned to their nearest centroid based on a distance measure (mostly Euclidean Distance Metric).

In the next step, the algorithm computes the new mean value of each cluster. To build this step the term "centroid update" cluster is used. Now that the centers are recalculated, each observation is once again tested to see whether it may be closer to a different cluster. All the objects are reassigned using the cluster updated means. The cluster assignment and centroid update steps are repeated iteratively till the cluster assignments cease to change (until convergence criterion is met).

The k-mediod/median was first introduced by two statisticians, Kaufman and Rousseeuw in 1987. It chooses data points as centers (medoid) to allocate the observations. It calculates in such a way that the total dissimilarity of all objects to their nearest medoid is minimal. It is said to be robust against k-means because it is capable of handling outliers or extreme values. The interpretations of the average were categorized into four groups [excellent/ strong structure (0.7- 1.0), very good structure (0.51-0.7), weak structure (0.26-0.50) and no substantial structure (<0.25) (Kaufman & Rousseeuw ,1987).

Ukekwe *et al.* (2023) conducted a study on clustering Internally Displaced Persons (IDP camps) for effective budgeting and resettlement policies using an optimized K-means approach, the research focused on the critical role of data-driven methods in humanitarian planning and resource allocation. It highlights prior research on clustering techniques for identified patterns in IDP distribution, camp characteristics, and resource needs. The study emphasizes the importance of clustering in grouping camps based on factors such as population size, geographical location, and infrastructure availability, which aids in prioritizing resources and formulating targeted policies. Using the demographic data published by displacement tracking matrix (DTM) on the IDP camps for 2021, an intelligent and optimized K-means cluster analysis was employed to classify the camps and identify their unique features for the purpose of guiding government policies and budget aimed at improving the

standard of living in the camps. In four clusters, the model successfully classified 164 of the camps as having major poverty and health issues, 114 as having overcrowding issues, and 13 as living a normal life. The result provides a suggestive guide for imminent aid.

Research was conducted on K-means cluster analysis of West African cereal species based on nutritional value composition would typically examine prior research on the use of clustering algorithms in agricultural and nutritional sciences. Atsa'am *et al.* (2021) the study highlights the importance of cereals like maize, millet, sorghum, and rice as staple foods in West Africa and their diverse nutritional profiles, which are essential for food security and dietary planning. Previous studies emphasize the application of K-means clustering in categorizing agricultural products based on nutrient composition, such as protein, carbohydrate, and mineral content, to identify patterns and groupings. The review might also discuss the role of clustering in crop improvement, market segmentation, and addressing malnutrition challenges. It explores how unsupervised learning techniques, like K-means, have been effectively employed to handle multidimensional nutritional datasets, offering insights into similarities and variations among cereal species for better agricultural planning and policy formulation.

Supriyatna *et al.,* (2020) conducted a research to classify rice production in 34 provinces in Indonesia in 2018. The data used were sourced from Statistics Indonesia. The method or approach used in the study is the K-Means cluster algorithm to classify rice productivity data by province in 2018. The results of the research are; (1) There are 19 provinces included in cluster 0 (Medium), (2) There are 4 provinces included in cluster 1 (High), and (3) There are 11 provinces that are included in cluster 2 (Low). Based on the results of the study, it was proven that there were 4 provinces in cluster 1 (High) they are West Java, Central Java, East Java and South Sulawesi, with the highest rice productivity. Three of them were on Java Island. It shows that Java still dominates the productivity of rice plants.

A study was conducted to classify soil data based on their fertility in the Northwest region of Nigeria. Data was obtained from the Department of Soil Science, Ahmadu Bello University, Zaria. Hayatu *et al.,* (2020), the data contain 400 soil samples containing 13 attributes. The relationship between soil attributes was observed based on the data. k-means clustering algorithm was employed in analyzing soil fertility clusters using soil attributes such as Nitrogen (N), Potassium (K), Phosphorus (P), Magnesium (Mg), Sodium (Na), Calcium (Ca), Organic carbon (OC), Electrical Conductivity (EC), Salinity(SL), Clay (CY), Sand (SN), Calcium chloride ($CaCl_2$) and pH contents. Results show that there is a positive relationship between pH and $CaCl_2$,Ca, Mg and EC and also a close negative relationship between SL, SN and CE. The remaining parameters are not related to one another.

Surya and Laurence (2019) employed clustering techniques to analyze an agricultural dataset using k-means and k-medoid algorithms. They classified the data and conducted a performance analysis comparing the two techniques based on specific performance metrics. Experimental analysis showed that k-means yielded lower accuracy and precision compared to k-medoid. The dataset used in this research was collected from an official agricultural government portal. It contains state-wise, district-wise, crop-wise, season-wise, and year-wise data on crop coverage and crop yield production in India. The study concluded that k-medoid performed better, with higher accuracy and precision than k-means. Based on these results, k-medoid proved to be a more effective clustering method for this agricultural dataset. This finding can assist future implementations in selecting the most appropriate clustering technique for similar datasets.

Harikumar and Surya (2015) conduct a comprehensive study on the application of K-Medoid clustering for heterogeneous datasets. The research highlighted the strengths of the K-Medoid algorithm in dealing with mixed-type data, its robustness to outliers, and its interpretability, making it a valuable tool for clustering in various domains. The findings contribute to the field by demonstrating the practical advantages of K-Medoid clustering and providing insights into its implementation and performance evaluation. Recent years have explored various clustering strategies to partition datasets comprising of heterogeneous domains or types such as categorical, numerical and binary. Clustering algorithms seek to identify homogeneous groups of objects based on the values of their attributes. These algorithms either assume the attributes to be of homogeneous types or are converted into homogeneous types. However, datasets with heterogeneous data types are common in real life applications, which if converted, can lead to loss of information. This paper proposes a new similarity measure in the form of triplet to find the distance between two data objects with heterogeneous attribute types. A new k-medoid type of clustering algorithm is proposed by leveraging the similarity measure in the form of a vector. The proposed K-medoid type of clustering algorithm is compared with traditional clustering algorithms, based on cluster validation using Purity Index and Davies Bouldin index. This study uses a data collected from a bank which has both numerical and categorical features as duration, job, marital status, education, and housing. Results show that the new clustering algorithm with new similarity measure outperforms k-means clustering for mixed datasets.

Mbuka and Anjaneyulu (2016), conducted research on application of k-means and k-mediod multivariate statistical methods for the purpose of revealing optimal clusters and assessing the consistency of individual districts within the group. Data used were extracted from united republic of Tanzania (Ministry of Agriculture, Livestock and Fisheries (MALF) and consist a total of 36 districts with both maize and beans yield. The study findings revealed that the 36 districts were grouped into six clusters using k-means algorithm. Using the k-mediod, it was revealed that only 11 districts were found to be very well structured since their silhouette width silhouette is above 0.5. Nevertheless, the clusters validation was done in such a way that individual district whose silhouette width is close to 1 was regarded as highly consistency clustered whereas districts with silhouette is greater or equal to 0.25 were said to be somehow well clustered and otherwise. The paper concluded that few districts that are very consistent given the threshold margin should be monitored and evaluated effectively to ascertain productivity

MATERIALS AND METHODS

Data Used in the Analysis

The National Agricultural Extension and Research Liaison Services (NAERLS), Ahmadu Bello University, Zaria, publication room provided the data for this study. The information includes estimates of rice yield in Nigeria Based on area, production, and productivity and it consist of 111 observations.

Euclidean Distance

It is the most commonly used in cluster analysis. This formula calculate the distance between two point i and k in a t-dimensional space by summing the squared difference of their

corresponding features and taking the square root. It can be expressed by the following equations:

$d(i_{r,}i_k) = \left[\sum_{j=1}^{t}(X_{rj} - X_{kj})^2\right]^{\frac{1}{2}}$ (1)

where,

$d(i,k)$ represent the Euclidean distance between two point I and k

X_{ij} is the coordinate of point i and a dataset

X_{kj} is the jth measurement of point k

t is the total number of features or dimension for each point in the dataset

$\sum_{j=1}^{t}$ summation symbol that you sum over all features from j = 1 to j = t

The K-Means Clustering Algorithm

The algorithm has the following steps:

Step 1:

We randomly select cluster (centroids); let's assume these are $C_{1,}C_{2,}C_{3,}...,C_k$ and we say that $C = \{C_{1,}C_{2,}C_{3,}.... C_k\}$where C is the set of all centroids.

Step 2:

In this step, we assign each data point to closed center; this is done by calculating the Euclidean distance: arg min dist.$\left(C_{i;}x\right)^2$ with$C_i \in C$.

where $C$is the set of centroids.

Step 3:

To determine the new centroid, we take the average of all the points that are allocated to that cluster.

$C_i = \frac{\sum_{x_i \epsilon s_i} x_i}{|S_i|}$ (2)

where$|S_i|$the collection of all points is allocated to the i^{th} cluster and $|C_i|$ is the number of point assigned to new cluster centroid is C_i.

Step: 4

In this step, we go through steps two and three again in this step until the cluster assignments remain the same. That means we keep running the algorithm till our cluster does not change.

The K-Medoid Clustering Algorithm

1. Set value of k (quantity of cluster) objects.
2. Pick randomly k objects from n objects as medoids.
3. Calculate the distance from each n objects to medoids objects with Euclidean Distance

$d(i_{r,}i_k) = \left[\sum_{j=1}^{t}\left(X_{rj} - X_{kj}\right)^2\right]^{\frac{1}{2}}$

4. S = current total cost – previous total cost
5. If S > 0, then the clustering process is stopped. But if S < 0, exchange non- medoid with medoid and Repeat until S>0

RESULTS AND DISCUSSION

Results of the K-means Clustering Algorithm

Based on the data analyzed using K-means and K-medoid with pre- defined k=6 clusters, the findings revealed that the crop yield for rice collected from 36 states of Nigeria and FCT were grouped into six clusters on the basis of nearest centroid. It was also indicated that states placed together produce similar rice yield.

Table 1: The K-means Clustering of Rice in 36 States of Nigeria and FCT

S/N	States	Cluster Description	Sil-width
1	Benue	4	0.41671284
2	FCT	2	0.39728170
3	Kogi	4	0.48547761
4	Kwara	2	0.25876657
5	Nasarawa	2	0.48064717
6	Niger	4	0.23604829
7	Plateau	3	0.37056614
8	Adamawa	3	0.54249583
9	Bauchi	3	0.64164572
10	Borno	3	0.42689336
11	Gombe	3	0.56489591
12	Taraba	2	0.27378658
13	Yobe	6	0.39817309
14	Jigawa	3	0.53427855
15	Kaduna	2	0.40587794
16	Kano	2	0.21979750
17	Katsina	3	0.54466373
18	Kebbi	4	0.27023445
19	Sokoto	6	0.53787308
20	Zamfara	6	0.40265098
21	Abia	5	0.41617056
22	Anambra	1	0.42094455
23	Ebonyi	6	0.22977767
24	Enugu	5	0.41251065
25	Imo	5	0.09943655
26	Akwaibom	1	0.34992360
27	Bayelsa	1	0.47295081
28	Cross river	6	0.51876549
29	Delta	5	0.23009497
30	Edo	6	0.09665177
31	Rivers	1	0.31007585

32	Ekiti	5	0.21095717
33	Ogun	5	0.58279656
34	Ondo	1	0.08921945
35	Osun	5	0.53073721
36	Oyo	5	0.51467562
37	Lagos	5	0.54113041

Studying Table 1 critically, one can deduce that for the rice yields; were classify into six clusters for the yields in Nigeria and FCT.

Cluster 1: This cluster has an average silhouette width of 0.33 and includes five states; Anambra, Akwa-Ibom, Bayelsa, Rivers and Ondo states. The first four states with (s_i) of 0.42094455, 0.34992360, 0.47295081, and 0.31007585 are somehow well clustered. While Ondo with 0.08921945 indicated as bad clustered.

Cluster 2: This cluster has an average silhouette width of 0.34 and includes six states; FCT, Kwara, Nassarawa, Taraba, Kaduna and Kano states. The first five states has a silhouette width of 0.39728170, 0.25876657, 0.48064717, 0.27378658 and 0.40587794, respectively which implies that the states are well clustered in terms of cereals yields; while Kano with 0.21979750 is regarded as bad cluster.

Cluster 3: This cluster has an average silhouette width of 0.52 and it's regarded as good cluster than clusters 1 and 2 above, it has seven states; Plateau, Adamawa, Bauchi, Borno, Gombe, Jigawa, and Katsina states. Plateau, and Borno has silhouette width of with 0.37056614, and 0.42689336 respectively indicating that they are somehow well clustered, while Adamawa, Bauchi, Gombe, Jigawa and Katsina states are very good clustered with silhouette width of 0.54249583, 0.64164572, 0.56489591, 0.53427855, and 0.54466373 respectively.

Cluster 4: This cluster has an averages silhouette width of 0.35 and it includes four states; Benue, Kogi, Niger, and Kebbi states. Benue, Kogi and Kebbi states has silhouette width of 0.41671284, 0.48547761, and 0.27023445 are somehow well clustered while Niger with 0.23604829 is regarded as bad clustered.

Cluster 5: This cluster has an averages silhouette width of 0.39 and it has nine states; Abia, Enugu, Imo, Delta, Ekiti, Ogun, Osun, Oyo and Lagos states. Abia and Enugu has silhouette width of 0.41617056, 0.41251065 respectively and are regarded as somehow well clustered, while Imo, Delta and Ekiti states are regarded as bad cluster with silhouette value 0.09943655, 0.23009497 and 0.21095717 < 0.25. Ogun, Osun, Oyo, and Lagos states with 0.58279656, 0.53073721, 0.51467562, 0.54113041 are very good clustered.

Cluster 6: This cluster has an averages silhouette width of 0.36 and it has six states; Yobe, Sokoto, Zamfara, Ebonyi, Cross river and Edo states. Yobe, and Zamfara with silhouette width of 0.39817309, and 0.40265098 are somehow well clustered, while Sokoto, Crossrivers with 0.53787308, 0.51876549, are declared as very good clustered. Ebonyi and Edo states are bad clustered with 0.22977767, 0.09665177.

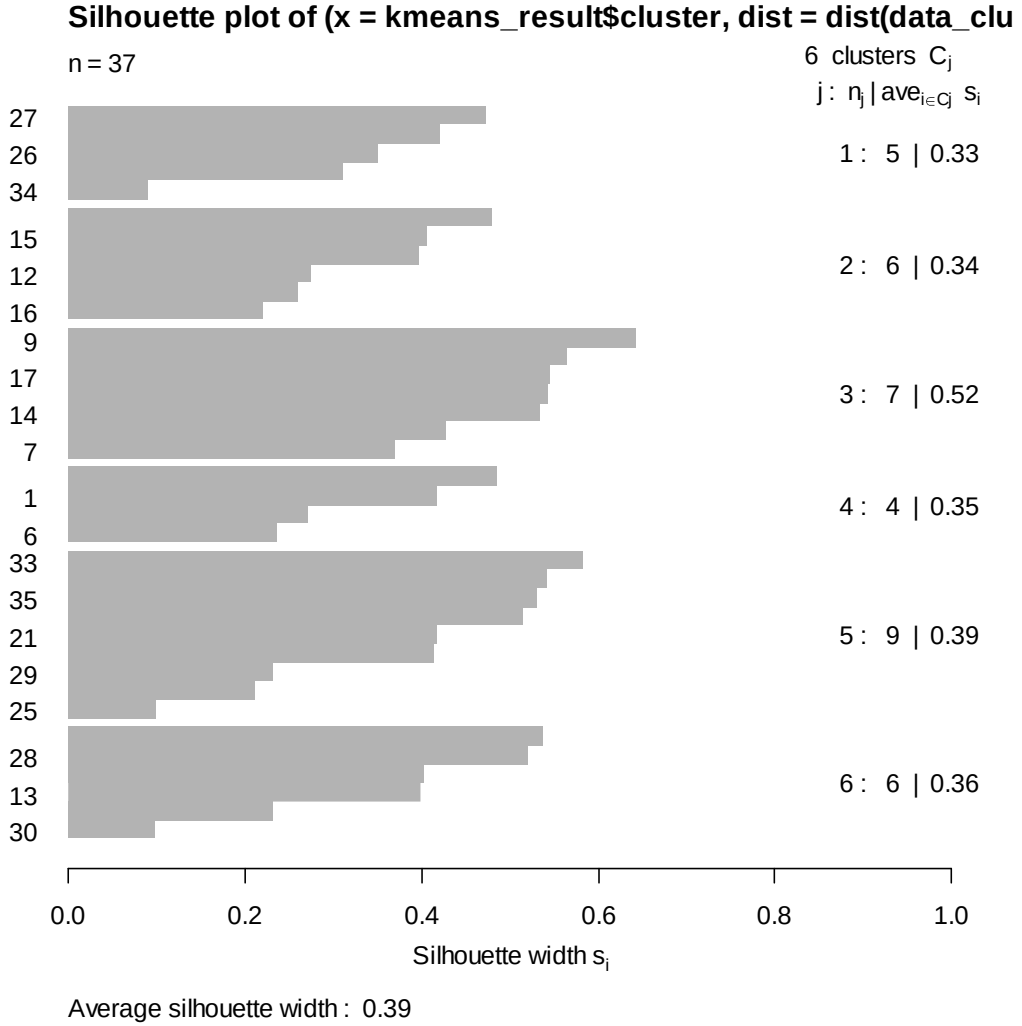

Figure 1: Average Silhouette Width Plot of K-means Clustering for Rice in 36 States of Nigeria and FCT

With reference to the figure 1 the number of elements (n_j) per cluster has been indicated. Each horizontal line corresponds to an element/state. The length of the lines corresponds to the silhouette width (s_i), which is the means similarity of each element to its own cluster minus the mean similarity to the next similar cluster. Also, it indicates the overall average silhouette width for six clusters validity. The average silhouette width for this clustering is 0.39, which suggests that the data points have been assigned to clusters somewhat well.

Table 2: Summary of K-means Clustering for Rice Yields in Nigerian States and FCT

	Rice		
Clusters Description	**No of element/state**	**Silhouette width**	**Average Silhouette width**
1	5	0.33	
2	6	0.34	
3	7	0.52	
4	4	0.35	0.39
5	9	0.39	
6	6	0.36	
No of Good Allocation	29		
No of Bad Allocation	8		

Table 2 gives the summary of k-means clustering applied to rice yields across six clusters. The number of element/states range from 4 to 9, with silhouette width varying from 0.33 to 0.52, indicating moderate good clustering quality. The average silhouette width for rice is 0.39 suggesting that the clustering performance is acceptable. The table shows 29 states correctly clustered while 8 are wrongly/ poorly clustered.

Result of K-Medoid Clustering Algorithm

Table 3: The K-Medoid Clustering of Rice in 37 Districts in Nigeria

S/N	States	Cluster Descript	Neighbor	Sil-width
1	Benue	1	2	0.27380524
2	FCT	2	1	0.70657777
3	Kogi	1	2	0.36377800
4	Kwara	2	1	0.60632874
5	Nassarawa	2	1	0.71368866
6	Niger	1	2	0.17067019
7	Plateau	3	4	0.44503598
8	Adamawa	3	2	0.41826367
9	Bauchi	3	2	0.61777602
10	Borno	3	4	0.48338548
11	Gombe	3	2	0.55760888
12	Taraba	2	1	0.60324245
13	Yobe	4	6	0.31808191
14	Jigawa	3	4	0.58204690
15	Kaduna	2	5	0.46964022
16	Kano	5	2	0.00000000
17	Katsina	3	4	0.59367212
18	Kebbi	1	2	0.07853936
19	Sokoto	4	6	0.51011925
20	Zamfara	4	3	0.26820304
21	Abia	6	4	0.39669446
22	Anambra	4	6	0.29250261
23	Ebonyi	4	6	0.58759427
24	Enugu	6	4	0.48679870
25	Imo	6	4	0.43807678
26	Akwaibom	6	4	0.16128202
27	Bayelsa	6	4	-0.10031013
28	Cross river	4	6	0.59063676
29	Delta	6	4	0.48670758
30	Edo	4	6	0.57883125
31	Rivers	6	4	0.26206108
32	Ekiti	6	4	0.14245944
33	Ogun	6	4	0.50954759
34	Ondo	4	6	0.47960121
35	Osun	6	4	0.41171069
36	Oyo	6	4	0.44150653
37	Lagos	6	4	0.54726773

Table 3 reveals that there are 6 clusters for the k medoid algorithm, where Cluster 1: has an average silhouette width (s_i) of 0.22 and includes four states: Benue, Kogi, Niger and kebbi states. These states are not properly clustered since their silhouette value is below 0.25. While Benue and Kogi states with (s_i) of 0.27380524 and 0.36377800 were found to be somehow well clustered, Kebbi and Niger with (s_i) of 0.07853936 and 0.17067019 <0.25 are closed to zero; thus they hold intermediate position between cluster 1 and 2. In other word they are not recommended as good clustered.

Cluster 2: has an (s_i) of 0.62 and includes five states: FCT, Kwara, Kogi, Nassarawa, Taraba, Kaduna states making it very good cluster compared to cluster 1 since their silhouette values is falls between 0.51-0.7. However, FCT and Nassarawa states are excellent clustered with (s_i) of 0.70657777, and 0.71368866. Kwara and Taraba with 0.60632874 and 0.60324245 are regarded as good clustered. While Kaduna with 0.46964022 indicates to be relatively well clustered.

Cluster 3: has an (s_i) of 0.53 and it consist of seven states: Plateau, Adamawa, Bauchi, Borno, Gombe, Jigawa, and Katsina states. This cluster is also a very good cluster since its silhouette values falls between 0.51-0.7. Plateau, Adamawa and Borno states with (s_i) of 0.44503598, 0.41826367 and 0.48338548, found to be somehow well clustered since their silhouttee falls between 0.26-0.50. Gombe, Jigawa, Katsina, and Bauchi states were good clustered with (s_i) of 0.55760888, 0.58204690, 0.59367212, 0.61777602.

Cluster 4: has an (s_i) of 0.45 and consist of eight states: Yobe, Sokoto, Zamfara, Anambra, Ebonyi, Cross river, Edo, and Ondo states. The narrow silhouette indicates a relatively weak cluster. However Yobe, Zamfara, Anambra, and Ondo states are declared as somehow well clustered with (s_i)of 0.31808191, 0.26820304, 0.29250261, and 0.47960121. Sokoto, Ebonyi, Cross river, and Edo states are good clustered with (s_i) of 0.51011925, 0.58759427, 0.59063676, 0.57883125 all between the range of 0.51- 0.7.

Cluster 5: has an (s_i)of 0.00 and consist of only one state Kano. This state is neither positioned to cluster 5 nor cluster 2. It is not recommended as a good cluster.

Cluster 6: has an (s_i)of 0.35 and consist of 12 states: Abia, Enugu, Imo, Akwa-Ibom, Bayelsa, Delta, Rivers, Ekiti, Ogun, Osun, Oyo, and Lagos states. This cluster is regarded as weak cluster. However, Abia, Enugu, Imo, Delta, Osun, Oyo, with

(s_i) of 0.39669446, 0.48679870, 0.43807678, 0.48670758, 0.41171069 and 0.44150653 Ogun and Lagos were also a good clustered with (s_i)of 0.50954759, 0.54726773. Akwa ibom and Ekiti states are not properly clustered. Bayelsa with -0.10031013 this depicted that, it is at an intermediate lying far from both cluster six and four. In other words it does not belong to any of the two clusters and it referred as bad clustered.

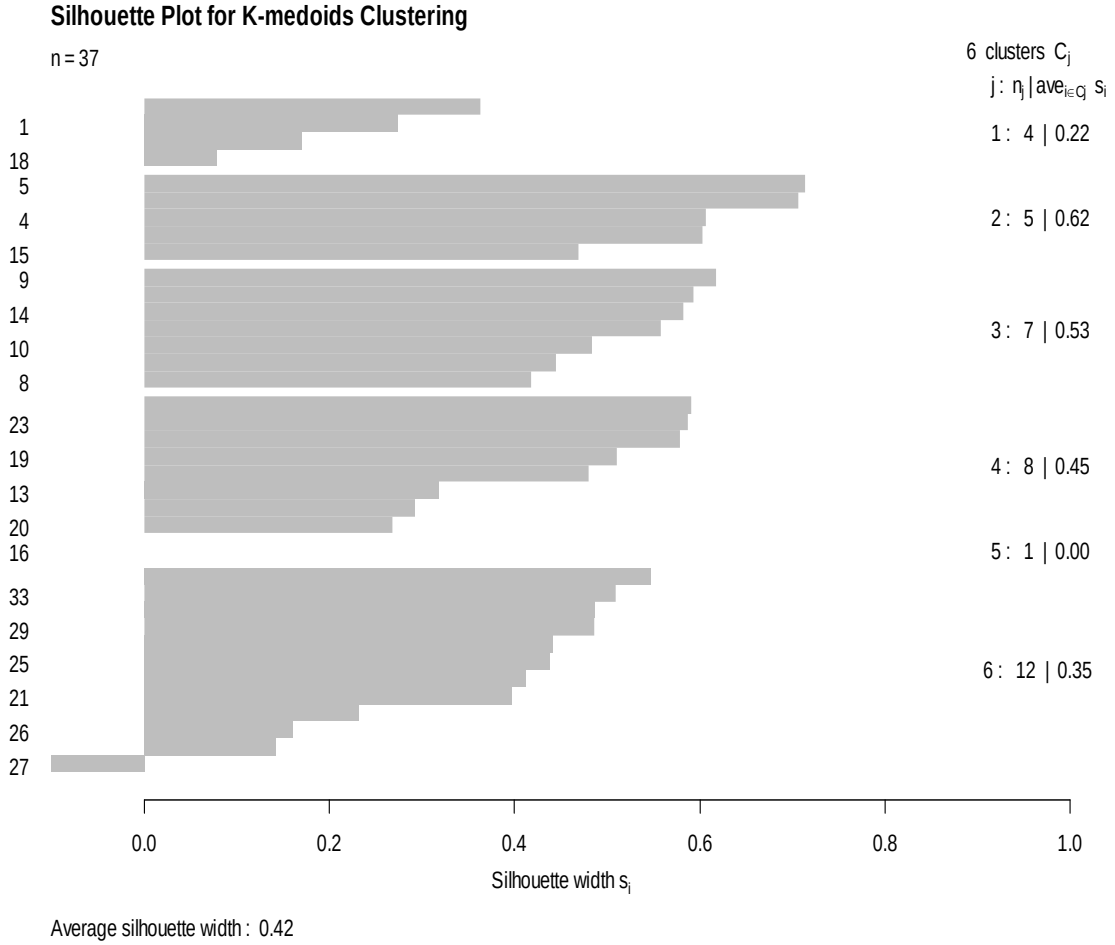

Figure 2: The Average Silhouette Width Plot of K-medoid Clustering for Rice in 36 States of Nigeria and FCT

In figure 2 the most common silhouette width in this graph is 0.35, which means a significant portion of data points are somewhat well-clustered. There are also data points with silhouette widths as high as 0.8, which means they are very well-clustered in their assigned groups. However, there are also data points with negative silhouette widths, which mean they may be better assigned to different clusters. The average silhouette width is 0.42, which again suggests a somewhat well-clustered dataset. Overall, the silhouette plot indicates that the k-medoid clustering resulted in a decent grouping of the data points.

Table 4: Summary of K-medoid Clustering for Rice Yields in Nigerian States and FCT

	Rice		
Clusters Description	**No of element/states**	**Silhouette width**	**Average Silhouette width**
1	4	0.22	
2	5	0.62	
3	7	0.53	0.42
4	8	0.45	
5	1	0.00	
6	12	0.35	
No of Good Allocation	31		
No of Bad Allocation	6		

Table 4 provides a summary of k-medoid clustering applied to rice yields, in Nigerian states and FCT indicating the clusters performance. Six clusters were identified, with a total of 31 elements (states) classified as good classification, and 6 as bad classification, the silhouette width range from 0.00 to 0.62, with an average silhouette width of 0.42, indicating moderate clustering performance. Both the clusters exhibit comparable clustering quality, with slight variations in individual cluster performance, as indicated by the silhouette width.

Table 5: Overall Summary of k-means and k-medoid clustering of rice showing good and bad Classification

Classifications	K-Means	K-Medoid
No of Good Allocation/Classification	29	31
No of Bad Allocation/Classification	8	6

Table 5 gives the overall summary of good and bad allocation for both the two clustering techniques (k-means & k-medoid) in which k-means has 29 and 8 good and bad classification respectively. While k-medoid has 31 and 6, good and bad classification respectively; which shows that k-medoid is the best techniques for classification than k-means. This finding is consistent with Kaufman and Rousseeuw. (1987), and Surya and Laurence (2019).

CONCLUSION

The analysis was carried out to evaluate the performance of two clustering techniques, k-means and k-medoid, clustering using agricultural data. The results of the two techniques ended with six clusters. The distribution of the observations among the clusters shows that cluster five has the highest observations using k- means with nine numbers of observations. The output of the k-medoid six clusters were also found and the silhouette measure of cohesion and separation chart (cluster quality) indicates that the overall model quality is fair than k-means with the value of silhouette 0.42 respectively. The k-means clustering has (29 & 8), number of good and bad allocations, while k-medoid has (31 & 6) number of good and bad allocations, from these values we observed that both methods fit our dataset, but the best fit is k-medoid clustering methods since it has the highest silhouette width than k-means and more clusters with Silhouette width closed to one and also has the highest number of good classification and a smaller number of bad classifications. The study recommends that government should pay attention on allocating the scarce resources to the consistency clusters along with policy review in favor of smallholder farmers through access and timely for all important farm inputs in future.

On the basis of these results, this paper recommends that when Nigerian government is planning to raise crop productivity, priority focus and attention should be on the clusters whose silhouette width (s_i) close to one since they are the best.

Again this paper recommends that the issue raising crop productivity should be in collaborations with various stakeholders such as individual farmers, government and non-government agencies, policy makers, planning units and seed manufacturing sfirms.

Statisticians are to be involved in data management for proper policy formulation;

More and expanded research is necessary in the agriculture for better understanding and proper management.

REFERENCES

Atsa'am, D. D., Oyelere, S. S., Balogun, O. S., Wario, R., & Blamah, N. V. (2021). K-means Cluster Analysis of the West African species of Cereals Based on Nutritional Value Composition. *African Journal of Food, Agriculture, Nutrition and Development*, *21*(1), 17195-17212.

Harikumar, S., & Surya, P. V. (2015). K-medoid Clustering for Heterogeneous Datasets. *Procedia Computer Science*, 70, 226-237.

Hayatu, I. H., Mohammed, A., Ismaâ, B. A., & Ali, S. Y. (2020). K-means clustering algorithm based classification of soil fertility in North West Nigeria. *FUDMA Journal of Sciences*, *4*(2), 780-787.

Kaufmann, & Rousseeuw. (1987). Clustering by Means of Medoids. In *Proc. Statistical Data Analysis Based on the L1 Norm Conference, Neuchatel, 1987* (pp. 405-416).

MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 281-297.

Mbukwa, J. N., & Anjaneyulu, G. (2016). Application of K-means and Partitioning Around Mediods (PAM) Clustering Techniques on Maize and Beans Yield in Tanzania. *Bulletin of Mathematics and Statistics Research*, 4(4), 146-158.

Supriyatna, A., Carolina, I., Widiati, W., & Nuraeni, C. (2020). Rice productivity analysis by Province Using K-means Cluster Algorithm. In *IOP Conference Series: Materials Science and Engineering*, 771(1), 12-25.

Surya, P. & Laurence, A. I. (2019). Performance Analysis of K-Means and K-Medoid Clustering Algorithms Using Agriculture Dataset. *Journal of Emerging Technologies and Iinnovative research*, 6(1), 1-7.

Ukekwe, E. C., Ogbonna, G. U. G., Adegoke, F. O., Okereke, G. E., & Asogwa, C. N. (2023). Clustering Nigeria's IDP Camps for Effective Budgeting and Re-Settlement Policies Using an Optimized K-Means Approach. *African Conflict & Peacebuilding Review*, 13(2), 60-85.

Wierzchoń, S. T., & Kłopotek, M. A. (2018). *Modern Algorithms of Cluster Analysis* (Vol. 34). Springer International Publishing.