FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 13, July, 2026, pp 204 – 210
DOI: https://doi.org/10.33003/fjs-2026-1013-5670

Comparative Analysis of K-Means and DBSCAN Clustering Techniques for Rice Yield Classification in Nigeria

***[1]Abduljalal Kabiru Hassan, [2]Nuraddeen Yusuf Ismail and [3]Ibrahim Muhammad Bunza**

[1]Department of Mathematics and Statistics, Air Force Institute of Technology, Kaduna, Nigeria.
[2]Department of Statistics, Ahmadu Bello University, Zaria, Kaduna, Nigeria.
[3]Department of Finance, Nigerian National Petroleum Company Limited, Nigeria.

*Corresponding authors' email: kachalla007@gmail.com

Received: 03 Jun. 2026
Accepted: 31 Jul. 2026
Published: 14 Aug. 2026

ABSTRACT
Rice production plays a vital role in ensuring food security and promoting agricultural development in Nigeria. This study compared the performance of the K-Means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithms in classifying rice yields across the 36 states of Nigeria and the Federal Capital Territory (FCT). Rice yield data for 2023 were obtained from the National Agricultural Extension and Research Liaison Services (NAERLS), Ahmadu Bello University, Zaria. K-Means partitioned the data into six clusters with an overall average silhouette width of 0.39, whereas DBSCAN identified three clusters and six noise points. Comparative analysis showed that DBSCAN achieved better clustering performance, with 31 allocated states compared with 29 for K-Means. The clustering patterns revealed regional similarities in rice yields, providing valuable insights for targeted agricultural planning and resource allocation. The findings demonstrate that DBSCAN is more effective than K-Means for clustering heterogeneous rice yield data, offering a reliable framework for evidence based agricultural planning and policy formulation in Nigeria.

Keywords: Cluster Analysis, Euclidean Distance, K-Means Clustering, DBSCAN Clustering

INTRODUCTION

Rice (Oryza sativa) is one of the world's most important cereal crops and a major staple food for more than half of the global population. Rice productivity varies across locations because of differences in soil properties, climate, and water availability, highlighting the need for location specific planning and management strategies (Usman *et al.,* 2022). In Nigeria, rice plays a critical role in food security, employment generation, poverty reduction, and economic development (Diagi *et al.,* 2021). The increasing demand for rice, driven by rapid population growth, urbanization, and changing dietary preferences, has intensified the need to improve domestic rice production and reduce dependence on imports. Consequently, the Nigerian government has implemented several agricultural intervention programmes aimed at increasing rice production and achieving self-sufficiency. Despite these efforts, considerable variations in rice yields continue to exist across the country's states due to differences in climatic conditions, soil characteristics, farming practices, access to improved inputs, irrigation facilities, and government support (Gama *et al.,* 2022).

Understanding the spatial distribution and regional variation in rice yields is essential for designing effective agricultural policies and allocating resources efficiently. States with similar production characteristics may require comparable intervention strategies, while regions with low productivity may benefit from targeted investments in improved seed varieties, mechanization, irrigation infrastructure, and extension services. Therefore, identifying homogeneous groups of rice producing states can provide valuable information for evidence based agricultural planning, monitoring, and decision making (Musa *et al.,* 2024)

Traditional statistical techniques are useful for summarizing agricultural data but usually fail to reveal hidden structures and natural groupings within complex datasets. Cluster analysis is an unsupervised machine learning technique that identifies groups of observations with similar characteristics without requiring prior class labels (Jaeger & Banks, 2023). By partitioning observations into relatively homogeneous clusters, cluster analysis facilitates the identification of underlying patterns and relationships that may not be apparent through conventional descriptive statistics (Jaeger & Banks, 2023). In agricultural research, clustering techniques have been successfully applied to classify crop production systems, evaluate crop performance, identify regional similarities, and support agricultural planning.

Several clustering algorithms have been applied across different disciplines, including agriculture, environmental science, business, healthcare, and spatial analysis. Previous studies have demonstrated the usefulness of Density Based Spatial Clustering of Applications with Noise (DBSCAN) and K-Means in analyzing complex datasets. For example, Atsa'am *et al.,* (2021) applied K-Means clustering to classify the nutritional composition of West African cereal crops, while Alqorni *et al.,* (2021) employed DBSCAN to classify citrus hybrids based on quality characteristics. Similarly, Fauzan *et al.,* (2022) used DBSCAN to identify spatial patterns in hotel distribution, and Sutramiani *et al.,* (2024) reported that DBSCAN outperformed K-Means in clustering micro, small, and medium enterprises because of its ability to detect clusters of varying densities and identify noise. These studies demonstrate the flexibility of clustering techniques across diverse application areas.

Despite the growing application of clustering algorithms, limited attention has been given to comparing centroid-based and density-based clustering techniques for classifying rice yields across Nigerian states. Most existing agricultural studies have focused on crop production trends, yield forecasting, or crop quality assessment, with relatively few studies exploring the identification of natural groupings among rice producing states using unsupervised machine learning techniques. Consequently, there is limited empirical

evidence regarding the most suitable clustering algorithm for analyzing heterogeneous rice yield data in Nigeria.

This study addresses this gap by comparing the performance of K-Means and DBSCAN in clustering rice yields across the 36 states of Nigeria and the Federal Capital Territory (FCT). Specifically, the study aims to identify natural groupings of rice producing states and evaluate the effectiveness of the two clustering algorithms using silhouette width as the primary performance measure. The findings are expected to provide useful information for policymakers, agricultural planners, researchers, and other stakeholders by identifying states with similar production characteristics and supporting more efficient allocation of agricultural resources and interventions. Furthermore, the study contributes to the growing application of unsupervised machine learning techniques in agricultural research by providing empirical evidence on the suitability of K-Means and DBSCAN for clustering heterogeneous agricultural datasets in Nigeria.

MATERIALS AND METHODS

Data Source

This study utilized secondary data on rice yields across the 36 states of Nigeria and the Federal Capital Territory (FCT). The data were obtained from the National Agricultural Extension and Research Liaison Services (NAERLS), Ahmadu Bello University, Zaria, for the year 2023. The dataset comprised estimates of rice production based on cultivated area, total of 111 observations. The data were used to evaluate the performance of K-Means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) in identifying homogenous groups of rice producing states.

Data Processing

Prior to clustering, the dataset was examined for completeness, consistency, and duplicate records. The rice yield variables were standardized using the z-score transformation to ensure that all variables contributed equally to the clustering process and to eliminate the influence of differences in measurement scales. Standardization was performed as:

$$z_i = \frac{x_i - \bar{x}}{s} \quad (1)$$

Where:
x_i =The original observation;
$\bar{x}$ =The sample mean; and
s =The sample standard deviation.

Euclidean Distance

The Euclidean distance was adopted as the similarity measure for clustering because it is one of the most widely used distance metrics in multivariate analysis. It measures the straight line distance between two observations in a multidimensional space and is expressed as:

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

Where:
$d(i,k)$ = represent the euclidean distance between observations (i) and (k);
X_{ij}= value of the j[th] variable for observation (i);
X_{kj}=value of the j[th] variable for observation (k);
t = total number of variables;
$\sum_{j=1}^{t}$ =summation over all j[th] variables from j = 1 to j = t.

K-Means Clustering Algorithm

The K-Means clustering algorithm, originally proposed by MacQueen (1967), partitions a dataset into a predetermined number of clusters (k) such that observations within the same cluster are as similar as possible while observations in different clusters are as distinct as possible. The K-means algorithm was implemented as follows:

i. Six initial centroids were randomly selected.
ii. Each observation was assigned to the nearest centroid using the Euclidean distance.
iii. The centroid of each cluster was recalculated as the mean of all observations assigned to that cluster.
iv. Steps 2 and 3 were repeated until no further changes occurred in the cluster assignments.

The centroid of cluster (k) is given by:

$$\mu_k = \frac{1}{n_k}\sum_{i=1}^{n_k} x_i \quad (3)$$

Where:
μ = Centroid of cluster (k);
n_k =Number of observations in cluster (k); and
x_i =Observation assigned to cluster (k).

Parameter Selection

The optimal number of clusters for the K-Means algorithm was determined using the Elbow Method. The within cluster sum of squares (WCSS) was computed for different values of k, and the point at which the rate of decrease in WCSS began to level off (the elbow) was selected as the optimal number of clusters. Based on this criterion, k = 6 was adopted for the analysis.

Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN is a density-based clustering algorithm that groups observations according to the density of neighbouring points rather than their distance from a centroid (Sutramiani *et al.,* 2024). Unlike K-Means, DBSCAN can identify clusters of arbitrary shapes and classify isolated observations as noise. The algorithm requires two parameters:

i. Epsilon (ε), which defines the radius of the neighbour hood around each observation.
ii. MinPts, which specifies the minimum number of neighbouring observations required for a point to be considered a core point.

In this study, the optimal parameter values were selected as $(\varepsilon = 0.72)$ and MinPts = 3, based on the k-nearest neighbour (K-NN) distance plot. The DBSCAN algorithm proceeds as follows:

i. Select an unvisited observation.
ii. Identify all neighbouring observations within the radius (ε).
iii. If the number of neighbouring observations is at least MinPts, the observation becomes a core point and a new cluster is formed.
iv. Expand the cluster by recursively including all density reachable observations.
v. Observations that do not satisfy the density requirement are classified as noise.
vi. Repeat the procedure until all observations have been visited.

The DBSCAN ε -Neighbourhood and the core point condition are given by:

$$N_\varepsilon(p) = \{q \in D : d(p,q) \leq \varepsilon\} \quad (4)$$

Where:
$N_\varepsilon(p)$ = Epsilon neighbourhood of point p;
D =Dataset;
$d(p,q)$ = Euclidean distance between point p and q;
ε =Neighbourhood radius.

$$|N_\varepsilon(p)| \geq MinPts \quad (5)$$

Model Evaluation

The quality of the clustering results was assessed using the silhouette coefficient, which measures the degree of similarity of an observation to its own cluster relative to neighbouring clusters. Silhouette values range from -1 to 1, where values close to 1 indicated well clustered observations, values around 0 indicate overlapping clusters, and negative values indicate poor cluster assignment. The overall average silhouette width was used to compare the performance of the K-Means and DBSCAN algorithms. Moreover, the numbers of well allocated observations were compared to evaluate the effectiveness of both clustering methods. The silhouette width equation is given by:

$$S(i) = \frac{b(i)-a(i)}{max\{a(i),b(i)\}} \quad (6)$$

Where:

$S(i) =$ Silhouette coefficient of observation$i$;

$a(i) =$ Average distance between observation $i$and all other observations in the same cluster;

$b(i) =$ Minimum average distance between observation $i$and observations in the nearest neighbouring cluster.

Statistical Software

All statistical analysis and clustering procedures were carried out using R statistical software. The implementation of the clustering algorithms and graphical visualizations was carried out using R packages, including cluster for silhouette analysis, factoextra for cluster visualization and validation, and dbscan for implementing the DBSCAN algorithm.

RESULTS AND DISCUSSION

K-Means Clustering of Rice Yields across Nigeria

The k- means and DBSCAN clustering algorithms were applied to classify rice yields across the 36 states of Nigeria and the Federal Capital Territory (FCT). Using the Elbow Method, the optimal number of clusters was determined as six (k = 6). The clustering results are presented in Table 1.

Table 1: K-Means Clustering of Rice in 36 States of Nigeria Including FCT

S/N	States	Cluster Description	Silhouette Width
1	Benue	4	0.41671284
2	FCT	2	0.39728170
3	Kogi	4	0.48547761
4	Kwara	2	0.25876657
5	Nassarawa	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

Table 1 show that the 37 observations were partitioned into six distinct clusters based on similarities in rice yields. Cluster 5 contained the highest number of states (9), followed by Cluster 3 (7), Clusters 2 and 6 (6 each), Cluster 1 (5), and Cluster 4 (4).

The average silhouette widths of the clusters from 0.33 to 0.52, indicating varying levels of clustering quality. Cluster 3 recorded the highest average silhouette width (0.52), suggesting the strongest cluster cohesion and separation from neighbouring clusters. In contrast, Cluster 1 had the lowest average silhouette width (0.33), indicating comparatively weaker clustering performance.

At the state level, Bauchi recorded the highest silhouette width (0.6416), followed by Ogun (0.5828), Gombe (0.5649),

and Lagos (0.5447), indicating that these states were well assigned to their respective clusters. Conversely, Ondo (0.0892), Edo (0.0967), and Imo (0.0994) recorded the lowest silhouette widths, suggesting weaker cluster membership and possible overlap with neighbouring clusters.

Overall, the K-Means algorithm produced an average silhouette width of 0.39, indicating a moderate clustering structure for the rice yield dataset.

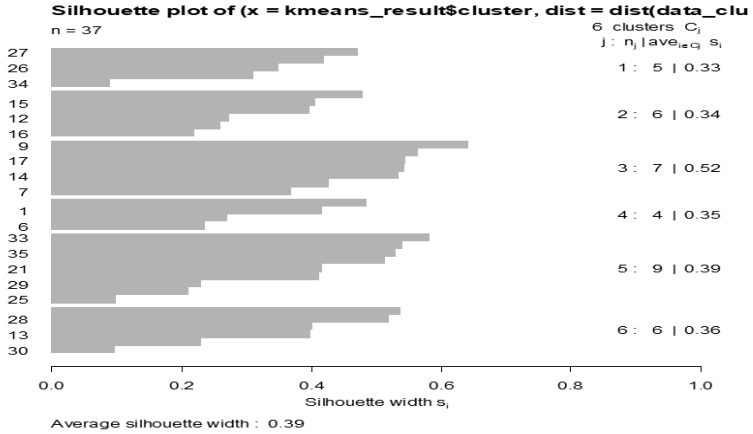

Figure 1: Silhouette Plot

Figure 1 presents the silhouette plot for the K-Means clustering solution. Each horizontal bar represents a state, while the length of the bar corresponds to its silhouette width. Positive silhouette values indicate that observations are appropriately assigned to their respective clusters, whereas values close to zero suggest overlapping cluster boundaries. The overall average silhouette width of 0.39 indicates that the six-cluster solution provides a moderately good representation of the underlying structure of the rice yield data.

Table 2: Summary of K-Means Clustering for the Rice Yields Shows in Table 1

	Rice		
Clusters Description	**No of State**	**Silhouette Width**	**Average Silhouette Width**
1	5	0.33	0.39
2	6	0.34	
3	7	0.52	
4	4	0.35	
5	9	0.39	
6	6	0.36	
Good Allocations	29		
Poor Allocations	8		

Table 2 summarizes the K-Means clustering results. The six clusters contained between four and nine states. Cluster 3 exhibited the highest average silhouette width (0.52), followed by Cluster 5 (0.39), while Cluster 1 recorded the lowest average silhouette width (0.33). Overall, the K-Means algorithm achieved an average silhouette width of 0.39, with 29 states classified appropriately and 8 states showing poor cluster allocation. These findings indicate that K-Means provided a reasonable clustering solution for the rice yield data.

The DBSCAN Clustering of Rice Yields in the 36 States of Nigeria and the FCT

Table 3: The DBSCAN Clustering of Rice in 36 States of Nigeria and FCT

S/N	**States**	**Cluster Description**	**Silhouette width**
1	Benue	0	-0.65122142
2	FCT	1	0.79637146
3	Kogi	2	0.49434488
4	Kwara	1	0.73953006
5	Nassarawa	1	0.71765428
6	Niger	0	-0.41757238
7	Plateau	3	0.14249548
8	Adamawa	3	-0.19635640
9	Bauchi	3	0.01676468
10	Borno	3	0.45719872
11	Gombe	0	-0.44045994
12	Taraba	1	0.61121216
13	Yobe	3	0.59139891
14	Jigawa	3	0.40653329
15	Kaduna	0	-0.58212231
16	Kano	0	-0.18134053

S/N	States	Cluster Description	Silhouette width
17	Katsina	3	0.37007240
18	Kebbi	2	0.50610743
19	Sokoto	3	0.57212513
20	Zamfara	3	0.46299617
21	Abia	0	-0.55244183
22	Anambra	3	0.58843281
23	Ebonyi	3	0.46624377
24	Enugu	3	0.67564611
25	Imo	3	0.67625132
26	AkwaIbom	3	0.59452733
27	Bayelsa	3	0.64806762
28	CrossRiver	3	0.52603996
29	Delta	3	0.62437819
30	Edo	3	0.49865240
31	Rivers	3	0.66358622
32	Ekiti	3	0.63539552
33	Ogun	3	0.60510852
34	Ondo	3	0.55033275
35	Osun	3	0.64915044
36	Oyo	3	0.66740543
37	Lagos	3	0.64742639

The DBSCAN algorithm was applied to the rice yield data using an epsilon (ε) value of 0.72 and MinPts = 3. Based on these parameter settings, the algorithm identified three clusters and six noise points, as presented in Table 3.

The clustering results showed that Cluster 3 was the largest, containing 25 states, followed by Cluster 1 with 4 states and Cluster 2 with 2 states. The remaining 6 states (Benue, Niger, Gombe, Kaduna, Kano, and Abia) were classified as noise because they did not satisfy the minimum density requirement for cluster membership.

Cluster 1 comprised the Federal Capital Territory (FCT), Kwara, Nasarawa, and Taraba, with silhouette widths ranging from 0.6112 to 0.7964, indicating strong cluster membership and clear separation from neighbouring clusters. Cluster 2 consisted of Kogi and Kebbi States, with silhouette widths of 0.4943 and 0.5061, respectively suggesting moderate clustering quality.

Cluster 3 contained the majority of the states, including Adamawa, Bauchi, Borno, Yobe, Jigawa, Katsina, Sokoto, Zamfara, Anambra, Ebonyi, Enugu, Imo, Akwa Ibom, Bayelsa, Cross River, Delta, Edo, Rivers, Ekiti, Ogun, Ondo, Osun, Oyo, and Lagos. Most states in this cluster recorded silhouette widths greater than 0.45, indicating good cluster cohesion. FCT recorded the highest silhouette width (0.7964), followed by Kwara (0.7394) and Nasarawa (0.7177), while Bauchi recorded the lowest positive silhouette width (0.0168), indicating weak but acceptable cluster membership. The six noise points recorded negative silhouette widths ranging from -0.6512 to -0.1813, indicating that these observations did not belong to any dense region and were appropriately identified as outliers by the DBSCAN algorithm.

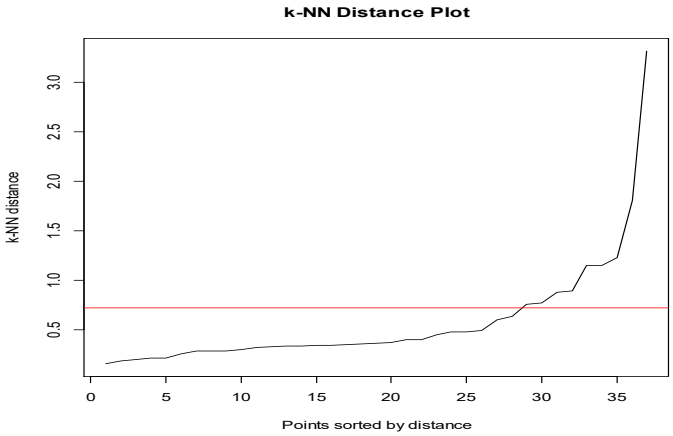

Figure 2: K-NN Distance Plot

Figure 2 presents the k-nearest neighbour (k-NN) distance plot used to determine the optimal epsilon (ε) value for the DBSCAN algorithm. The plot exhibits a noticeable change in slope at approximately 0.72, which was selected as the optimal epsilon value. This value provides an appropriate neighbourhood radius for identifying dense regions while minimizing the inclusion of isolated observations into clusters.

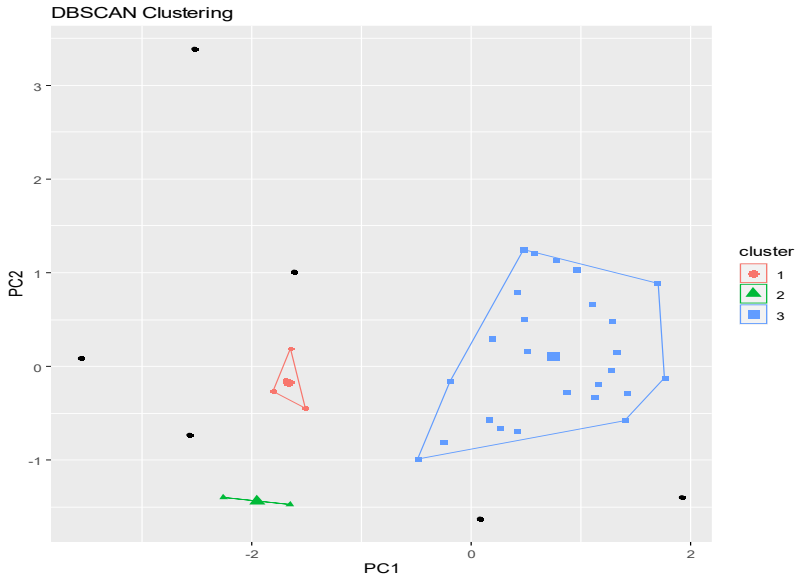

Figure 3: DBSCAN Cluster Plot

Figure 3 illustrates the DBSCAN clustering results based on the first two principal components. Different colours represent the identified clusters, while the black points denote observations classified as noise. Cluster 3 forms the largest and most compact group, whereas Clusters 1 and 2 contain relatively fewer states. The separation between the clusters demonstrates the ability of DBSCAN to identify density-based structures while simultaneously detecting observations that do not belong to any cluster.

Table 4: Comparison of K-Means and DBSCAN

Performance Metric	K-Means	DBSCAN
Number of Clusters	6	3
Good Allocations	29	31
Poor Allocations	8	6
Noise Points Detected	0	6

Table 4 presents a comparative evaluation of the K-Means and DBSCAN clustering algorithms for classifying rice yields across 36 states of Nigeria and the Federal Capital Territory (FCT). The K-Means algorithm partitioned the dataset into six clusters and correctly classified 29 states, with 8 states exhibiting poor cluster allocation. In contrast, DBSCAN identified three density-based clusters and six noise points, correctly classifying 31 states while recording only 6 poor allocations.

The comparative results indicate that DBSCAN achieved better clustering performance than K-Means by correctly classifying two additional states and reducing the number of poor allocations by two. Unlike K-Means, which assigns every observation to a cluster regardless of its similarity, DBSCAN is capable of identifying observations that do not belong to any dense region and classifying them as noise. This ability enables DBSCAN to produce more homogeneous clusters and improve the overall quality of the clustering solution for heterogeneous rice yield data. Overall, the findings suggest that DBSCAN is a more suitable clustering algorithm for rice yield classification because of its robustness in handling variations in the data and its ability to detect outliers while maintaining higher classification accuracy than the K-Means algorithm.

Discussion

The findings of this study demonstrate that both K-Means and DBSCAN were effective in identifying patterns in rice yields across the 36 states of Nigeria and the Federal Capital Territory (FCT). However, DBSCAN outperformed K-Means by correctly classifying more states (31 vs 29) and reducing the number of poorly classified states (6 vs 8). This superior performance is attributable to DBSCAN's ability to identify density-based clusters and detect outliers, making it more suitable for heterogeneous agricultural datasets where production patterns vary considerably across regions.

The results are consistent with the findings of Sutramiani *et al.,* (2024), who reported that DBSCAN produced more meaningful clusters than K-Means because of its ability to handle noise and clusters with irregular shapes. Similarly, Fauzan *et al.,* (2022) found that DBSCAN effectively identified spatial clusters while minimizing the influence of outliers, thereby improving clustering quality. The present study extends these findings to rice yield classification in Nigeria, demonstrating the applicability of density-based clustering techniques to agricultural production data.

The clustering patterns observed reflect regional similarities in rice production, suggesting that states within the same cluster share comparable production characteristics that may be influenced by factors such as agroecological conditions, rainfall distribution, soil fertility, farming practices, and access to agricultural inputs. Identifying these homogeneous groups provides valuable information for designing targeted agricultural interventions, improving resource allocation, and supporting evidence-based planning.

Despite these findings, this study has some limitations. The analysis was based on rice yield data for 2023 only and considered a limited number of production variables. Consequently, the identified clusters may not fully capture temporal variations in rice production across Nigeria. Future

studies should incorporate multi-year datasets and additional environmental, climatic, and socioeconomic variables to provide a more comprehensive assessment of spatial patterns in rice productivity.

CONCLUSION

This study evaluated the performance of the K-Means and DBSCAN clustering algorithms in classifying rice yields across the 36 states of Nigeria and the Federal Capital Territory (FCT). The findings demonstrated that both algorithms successfully identified meaningful patterns in the rice yield data; however, DBSCAN provided a more robust clustering solution because of its ability to detect outliers and identify density-based clusters. These findings underscore the importance of selecting appropriate clustering techniques for agricultural datasets with heterogeneous production characteristics.

The study contributes to the application of unsupervised machine learning techniques in agricultural data analysis by demonstrating the suitability of DBSCAN for regional rice yield classification in Nigeria. The resulting clusters provide useful information that can support evidence-based agricultural planning, targeted interventions, and efficient allocation of resources aimed at improving rice productivity and strengthening food security.

This study was limited to rice yield data for 2023 and considered only the available production variables. Future studies should incorporate multi-year datasets together with climatic, environmental, and socioeconomic variables to improve the robustness and generalizability of the clustering results.

Based on the findings of this study, the following recommendations are made:

i. Government agencies and agricultural planners should utilize clustering techniques, particularly DBSCAN, to identify states with similar rice production characteristics for targeted policy interventions and efficient resource allocation.
ii. Investments in agricultural infrastructure, improved seed varieties, extension services, and access to farm inputs should be tailored to the specific needs of the identified clusters to enhance rice productivity.
iii. Collaboration among researchers, policymakers, agricultural extension officers, and farmers should be strengthened to promote data driven decision making in agricultural planning and development.
iv. Future research should investigate the performance of additional clustering algorithms using multi-year rice production data and incorporate climatic, environmental, and socioeconomic factors to improve the accuracy and applicability of clustering outcomes.

REFERENCES

Alqorni, F., Mahmudy, W. F., & Widodo, A. W. (2021). Application of density-based spatial clustering of applications with noise (DBSCAN) in determining the quality of Keprok orange and Siam orange hybrids at the Citrus and Subtropical Crops Research Institute, Batu City. Journal of Information Technology and Computer Science, 6(1), 1–8. https://doi.org/10.25126/jitecs.202161244

Atsa'am, D. D., Oyelere, S. S., Balogun, O. S., Wario, R., & Blamah, N. V. (2021). K-means cluster analysis of West African cereal species based on nutritional value composition. African Journal of Food, Agriculture, Nutrition and Development, 21(1), 17195–17212.

Diagi, B. E., Edokpa, D. O., & Ajiere, S. I. (2021). Rice and climate change: Its significance towards achieving food security in Nigeria: A review. Journal of Geographical Research, 4(1), 18-27. https://doi.org/10.30564/jgr.v4i1.2588

Fauzan, A., Novianti, A., Ramadhani, R. R. M. A., & Adhiwibawa, M. A. S. (2022). Analysis of hotel spatial clustering in Bali: A density-based spatial clustering of applications with noise (DBSCAN) algorithm approach. EKSAKTA: Journal of Sciences and Data Analysis, 3(1), 25–38. https://doi.org/10.20885/EKSAKTA.vol3.iss1.art4

Gama, E., Mukhtar, U., & Choumbou, R. F. D. (2022). Resource-use efficiency of rice production in Kura Local Government Area of Kano State, Nigeria. FUDMA Journal of Agriculture and Agricultural Technology, 8(1), 131–141. https://doi.org/10.33003/jaat.2022.0801.036

Jaeger, A., & Banks, D. (2023). Cluster analysis: A modern statistical review. WIREs Computational Statistics, 15(3), Article e1597. https://doi.org/10.1002/wics.1597

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, pp. 281–297). University of California Press.

Musa, M. H., Agomuo, J. K., & Dandago, M. A. (2024). Paddy rice (Oryza sativa) production and processing in Nigeria: A review. Dutse Journal of Pure and Applied Sciences, 10(1b), 282-292. https://doi.org/10.4314/dujopas.v10i1b.28

Raj, C. K. (2017). Comparison of K-means, K-medoids, and DBSCAN algorithms using a DNA microarray dataset. International Journal of Computational and Applied Mathematics, 12(1), 344-355.

Sutramiani, N. P., Arthana, I. K. R., Aurelia, S., Fauzi, M., & Surya Darma, I. (2024). The performance comparison of DBSCAN and K-means clustering for MSMEs grouping based on asset value and turnover. Journal of Information Systems Engineering and Business Intelligence, 10(1), 13–24.

Usman, J., Usman, M., Ogbu, O. J., & Azagaku, E. D. (2022). Suitability evaluation of rainfed rice (Oryza sativa) in the floodplain of the Southern Guinea Savannah Zone of Nigeria. FUDMA Journal of Agriculture and Agricultural Technology, 8(1), 339–350. https://doi.org/10.33003/jaat.2022.0801.102

Wibisono, S., Anwar, M. T., Supriyanto, A., & Amin, I. H. A. (2021). Multivariate weather anomaly detection using the DBSCAN clustering algorithm. Journal of Physics: Conference Series, 1869(1), Article 012077.