Clustering Analysis of Subsidized Fertilizer Recipients in 2025 Using K-Means++ and Fuzzy C-Means

Umar Al Faruq[1], Rini Indriati[2], Aidina Ristyawan[3]
Universitas Nusantara PGRI Kediri, Jl. KH. Achmad Dahlan No. 76, 64112, Indonesia
E-mail: umarfaroq1991@gmail.com, rini.indriati@unpkediri.ac.id, adinaristi@unpkediri.ac.id
Corresponding Author: Umar Al Faruq

Abstract: *This study aims to analyze and compare the performance of clustering models in grouping subsidized fertilizer recipient data in 2025 to support the efficiency and accuracy of government distribution policy targets. The recipient data were processed through feature selection, data transformation, data type conversion, and missing value handling. The K-Means++ and Fuzzy C-Means (FCM) clustering methods are applied with the optimal number of clusters (K) set at eight (K=8) based on validity metric analysis. The model evaluation results show that the K-Means++ algorithm produces better cluster quality than FCM. The internal validity metric assessment for K-Means++ on the K=8 cluster shows a Silhouette Score of 0.756, a Davies-Bouldin Index (DBI) of 0.241, and a Calinski-Harabasz index (CHI) of 7901. Meanwhile, FCM reports S=0.737, DBI=0.424, and CHI=4903. This comparison clearly shows that K-Means++ has advantages in terms of clearer cluster separation and stability. The conclusion of this study is that the K-Means++ algorithm is the most effective model and is recommended for use in grouping recipients of subsidized fertilizer assistance in 2025. The results of this grouping (8 clusters) can present an accurate and useful profile for policymakers in designing more effective fertilizer allocation and distribution priority strategies.*

Key Words*: Clustering Analysis; Subsidized Fertilizer; K-Means++; Fuzzy C-Means*

Introduction

Fertilizer subsidies are a strategic government policy aimed at increasing agricultural productivity while maintaining national food security. Each year, the government allocates a significant budget to ensure the availability of subsidized fertilizer for farmers. However, the implementation of this policy still faces various obstacles, including inaccurate targeting of recipients, unequal distribution of fertilizer among farmers, and limited regular updates of subsidized fertilizer recipient data.(Ika Ari Sasmita et al., 2021;Hermantria, 2021). This problem has the potential to reduce the effectiveness of the fertilizer subsidy policy and cause inefficiency in the use of the state budget.

As the complexity of agricultural data management increases, the application of a data-driven approach is becoming increasingly important in supporting more objective and measurable policy formulation. In this context, data mining techniques play a crucial role in extracting hidden patterns from large-scale data, particularly through clustering methods that aim to group data based on the degree of similarity of certain characteristics. (Kodinariya, T.M & Makwana, P.R, 2021). Clustering techniques have been widely applied in the agricultural sector to systematically map land characteristics, production inputs, and farmer groups. (Aldino et al., 2021)

A number of previous studies have examined the application of clustering in the context of the fertilizer and agricultural sectors. Zalfa et al., (2023) applied the K-Means algorithm to cluster subsidized fertilizer recipients based on RDKK data. The results of this study indicate that the clustering method is capable of grouping farmers based on land area and the amount of fertilizer received, thus it can be used as a basis for consideration in the equitable distribution of subsidized fertilizer. However, this study is still limited to the use of a single clustering algorithm and a single evaluation indicator, so the quality and stability of the clustering results have not been comprehensively analyzed.

Another study conducted by Hermantria, (2021) examined the segmentation of non-subsidized fertilizer consumers using the K-Means method. The results showed that clustering was effective in identifying farmer groups with specific characteristics and provided useful information for strategic decision-making in the fertilizer sector. However, the study focused more on non-subsidized fertilizer and market segmentation aspects, so it did not specifically address the issue of subsidized fertilizer distribution.

In addition to algorithm selection, determining the optimal number of clusters is an important aspect of clustering analysis. Several studies have confirmed that using a single evaluation index is insufficient to guarantee the quality of clustering results (Meng et al., 2023). Silhouette Score is used to measure the level of cohesion and separation between clusters, Davies-Bouldin Index is used to assess the level of similarity between clusters, while Calinski-Harabasz Index is used to measure the ratio of variation between clusters and variation within clusters (Zhang, Liu & Wang, 2023). Similar findings were also shown in studies on agricultural databases which emphasized the importance of selecting initial cluster centers to avoid biased clustering results.

On the other hand, soft clustering approaches such as Fuzzy C-Means (FCM) allow a single data item to have membership in more than one cluster, thus better representing agricultural data that lacks clear boundaries between clusters. This approach is relevant in the context of subsidized fertilizer recipient data, given that farmer characteristics often overlap based on land area and fertilizer needs.

Based on the results of a review of previous studies, a research gap can be identified, namely the limited number of studies comparing more than one clustering method on subsidized fertilizer recipient data using multiple cluster validation indices simultaneously. Most previous studies only used one algorithm and one evaluation indicator and did not compare hard and soft clustering approaches on the latest subsidized fertilizer data.

Thus, the scientific novelty, the novelty in this study lies in the application of two different clustering methods, namely K-Means++ and Fuzzy C-Means, in grouping recipients of subsidized fertilizer in 2025. In addition, this study uses three cluster validation indices simultaneously, namely Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index, to determine the optimal number of clusters. This approach is expected to produce more accurate, stable, and easily interpreted groupings as a basis for decision-making in subsidized fertilizer distribution policies.

Based on the description, the objective of this study is to group recipients of subsidized fertilizer in 2025 using the K-Means++ and Fuzzy C-Means methods, determine the optimal number of clusters based on several cluster validation indices, and compare the performance of the two clustering methods in supporting decision-making on subsidized fertilizer distribution policies.

Method

1. Research Design and Approach

This study applies a quantitative, descriptive, and analytical approach to group subsidized fertilizer recipients based on specific numerical characteristics using clustering techniques. The quantitative approach is used because the analysis focuses on processing numerical data to identify patterns and structures objectively and measurably. Meanwhile, the descriptive and analytical approach plays a role in systematically presenting and interpreting the characteristics of the clustering results (Putu Gede Subhaktiyasa et al., 2025).

Clustering techniques fall under the category of unsupervised learning, which aims to discover natural patterns in data without requiring prior class labels. This approach is relevant for data on subsidized fertilizer recipients, which have heterogeneous characteristics and do not always show clear group boundaries. Therefore, clustering is performed based on the degree of similarity in numerical characteristics (Kodinariya & Makwana, 2021).

In agriculture, clustering methods have been widely used for land analysis, farmer segmentation, and objective mapping of production input needs. Research by Aldino et al., (2021) The results show that applying clustering to agricultural data can identify groups with similar characteristics and support data-driven decision-making. Therefore, the use of a quantitative, analytical, descriptive approach through clustering techniques is considered appropriate for producing objective and scientifically accountable groupings of subsidized fertilizer recipients.

2. Research Objects and Data

The research object is data on subsidized fertilizer recipients in 2025 obtained from secondary data sources on fertilizer distribution. This dataset represents fertilizer recipients with several quantitative variables relevant to the clustering process. The variables used in the analysis include agricultural land area, the amount of urea fertilizer, NPK fertilizer, and organic fertilizer. Identity variables were not included in the analysis process because they do not represent numerical similarity characteristics, as is common practice in *unsupervised learning* research on agricultural and fertilizer data.

3. Software and Analysis Tools

The analysis process was conducted using Python-based software run through Google Colaboratory. Several libraries were used to support the analysis, including pandas and numpy for data processing and manipulation, scikit-learn for implementing the K-Means++ algorithm and calculating the clustering validation index, and scikit-fuzzy for implementing the Fuzzy C-Means algorithm. The use of this software aligns with the *machine learning* analysis approach widely applied in clustering research in agriculture. (Prakash et al., 2022).

4. Data Pre-processing

The research data was then processed through a pre-processing stage before conducting clustering analysis. This stage was carried out to ensure the data used was of good quality and suitable for the clustering process. Pre-processing included selecting numeric variables relevant to the research objectives, checking the completeness of the data to identify missing or invalid values, and adjusting the data format to prepare it for further analysis.

Afterward, data normalization was performed using the Standard Scaler method to equalize the scales between variables. Normalization is necessary so that each variable has a balanced influence on the calculation of distance between data. This is important because distance-based clustering algorithms, such as K-Means++ and Fuzzy C-Means, are very sensitive to differences in scale. Without normalization, variables with a larger range of values can dominate the clustering results and result in the formation of less representative clusters. Therefore, applying normalization is a crucial step to obtain more stable and accurate clustering results.

5. Clustering K-Means++

The K-Means++ algorithm is implemented as a hard clustering method that assigns each data item of subsidized fertilizer recipients to only one group based on their proximity to the cluster center. This approach produces precise groupings, simplifying the interpretation of the results, especially when the analysis is used as a basis for policy formulation that requires clear and operational classification.

The advantage of K-Means++ over conventional K-Means lies in the more structured centroid initialization process. The initial cluster centers are not selected randomly but instead consider the distance distribution between the data points, resulting in a more evenly distributed distribution of the initial centroids. This mechanism reduces the dependence of clustering results on initial conditions and reduces the risk of unstable or biased clusters. (Daoudi et al., 2021).

Improvements in the initialization stage directly impact the quality of the clustering results, both in terms of cluster compactness and separation between clusters. Various studies of clustering algorithm development have shown that K-Means++ is often used as the basis for advanced optimization methods due to its ability to produce more consistent cluster structures, especially on large and heterogeneous datasets. (Kumar & Singh, 2022). In the context of agricultural data that has variations in land characteristics and fertilizer requirements, these characteristics are important to produce clusters that are more representative of actual conditions.

Therefore, the use of the K-Means++ algorithm in this study is expected to be able to produce a more stable, accurate, and easy-to-understand grouping of subsidized fertilizer recipients, so that the analysis results can be used effectively as a basis for supporting decision-making in subsidized fertilizer distribution policies.

6. Clustering Fuzzy C-Means

In addition to the hard clustering approach using K-Means++, this study also applied the Fuzzy C-Means (FCM) algorithm as a soft clustering method. Unlike K-Means++, FCM does not strictly assign data to a single cluster but rather assigns a membership degree to each data item across all the clusters formed. This approach allows a single subsidized fertilizer recipient to have varying degrees of proximity to multiple clusters, allowing for a more realistic representation of the data's ambiguous structure.

Conceptually, FCM works by minimizing an objective function that considers the distance between the data and the cluster centroid, as well as the membership weights of each data item. Each iteration of the algorithm updates the membership values and cluster centroids until convergence is achieved. This mechanism makes FCM more flexible in capturing overlapping data patterns, particularly in agricultural data, which is influenced by multiple factors and does not always have clear cluster boundaries. This aligns with the findings of Krasnov et al., (2023) which states that FCM is effective for data with heterogeneous characteristics because it can model uncertainty and ambiguity between clusters through fuzzy membership values.

In the context of subsidized fertilizer recipient data, this heterogeneity can arise from differences in land area, variations in fertilizer types used, and patterns of fertilizer needs that are not always uniform across farmers. Therefore, the use of FCM in this study aims to capture the phenomenon that some fertilizer recipients can fall between two characteristic groups, for example, farmers with medium-scale land holdings who have fertilizer need patterns close to those of the small and large clusters.

Recent research also shows that FCM performance is significantly influenced by the initialization process for cluster centers and the distance calculations used. Seyed Emadedin Hashemi et al., (2023) confirmed that conventional FCM has limitations in terms of sensitivity to initial initialization and potential convergence to local solutions. To address this, various

FCM developments have been conducted, including optimizing centroid initialization and modifying the distance function, which have been shown to improve the quality of clustering results and model stability. These findings reinforce the relevance of using FCM as a comparison to K-Means++ in clustering analysis, particularly to evaluate the extent to which soft clustering approaches can provide additional insights into data structure.

Nevertheless, the FCM clustering results in this study indicate that the quality of cluster separation numerically is still below that of K-Means++. This is reflected in the lower Silhouette Score and Calinski-Harabasz Index values, as well as the higher Davies-Bouldin Index values. This indicates that although FCM is able to capture overlap between clusters, the resulting cluster structure is not as compact and robust as the results from K-Means++. However, as explained by Krasnov et al. (2023), The advantage of FCM lies not solely in its sharp cluster separation, but rather in its ability to represent close relationships among data in a gradual manner, which is often more suitable for exploratory analysis.

Thus, the application of FCM in this study provides a complementary perspective to the results of K-Means++. While K-Means++ excels in producing clear and easily interpretable clusters for policy purposes, FCM contributes to understanding the complexity and overlapping characteristics of subsidized fertilizer recipients. The combination of these two approaches enriches the clustering analysis and strengthens the overall interpretation of the clustering results.

7. Determining the Optimal Number of Clusters

Determining the optimal number of clusters is a crucial step in clustering analysis because it directly impacts the quality and ease of interpretation of the clustering results. Selecting an inappropriate number of clusters can result in data with different characteristics being grouped together, or conversely, resulting in too many clusters that are difficult to interpret. Therefore, this study tested several clustering options to determine the clustering structure that best fits the data characteristics.

The quality of the clustering results was evaluated using three validation indices: the Silhouette Score, the Davies-Bouldin Index, and the Calinski-Harabasz Index. The Silhouette Score is used to assess the suitability of data to its assigned cluster by comparing the average distance of the data to its own cluster and to the nearest other cluster. A high Silhouette Score indicates that the data is in the correct cluster and has clear separation between clusters. The Davies-Bouldin Index measures the degree of closeness between clusters by considering the ratio of variation within clusters to the distance between clusters. A lower index value indicates that the formed clusters are more compact and have minimal overlap. Meanwhile, the Calinski-Harabasz Index assesses the balance between variation between clusters and variation within clusters. A higher index value indicates that differences between clusters are more dominant than variations within clusters.

The simultaneous use of these three validation indices aims to provide a more comprehensive assessment of clustering quality. By considering various evaluation aspects, such as cohesion, separation, and cluster variation structure, the determination of the optimal number of clusters in this study is conducted more objectively and does not rely solely on a single evaluation measure. This multi-index approach provides a strong basis for determining the most representative number of clusters to describe patterns in subsidized fertilizer recipient data.

8. Data Determination and Interpretation

The clustering results were then analyzed descriptively by calculating the average (mean) value for each variable in each cluster. This descriptive analysis was conducted to illustrate the dominant characteristics of each group of subsidized fertilizer recipients, such as differences in agricultural land area and fertilizer requirement patterns between clusters. By calculating the average value, each cluster can be quantitatively represented, allowing for clearer and more systematic observations of differences between groups.

This approach enabled researchers to identify distribution patterns and trends in fertilizer use within each cluster, such as recipient groups with relatively small land areas but high fertilizer use intensity, or vice versa. This information provided a more in-depth picture of the data structure of subsidized fertilizer recipients and helped explain the substantive meaning of the clustering results, beyond simply numerical grouping.

The clustering results were then analyzed descriptively by calculating the average value for each variable in each cluster. This analysis aimed to identify the characteristics of each group of subsidized fertilizer recipients so that the clustering results could be interpreted substantively and applied as a basis for supporting data-driven decision-making in subsidized fertilizer distribution policies, as demonstrated in clustering research in the agricultural sector. (Swain et al., 2024).

Results and Discussion

1. Determining the Optimal Number of Clusters

Determining the optimal number of clusters is performed as an initial step to ensure the quality of the clustering results. Evaluation was conducted using three validation indices: the Silhouette Score, the Davies-Bouldin Index, and the Calinski-Harabasz Index. A summary of the evaluation results for determining the optimal number of clusters is presented in Table 1.

Table 1. Evaluation Results of Determination of the Optimal Number of Clusters

Validation Index	K Optimal	Value
Silhoute Score	8	0,756
Davies-Bouldin Index	7-8	0,241
Calinski-Harabasz Index	8	7901

Based on Table 1 above, the highest Silhouette Score value obtained at K = 8 shows that in the number of clusters, the data of subsidized fertilizer recipients has a high level of cohesiveness in the cluster and a clear separation between clusters. This condition indicates that each fertilizer recipient has more characteristic proximity to the same cluster members compared to other clusters. On the other hand, the low values of the Davies-Bouldin Index at K = 7 and K = 8 indicate that the distance between clusters is relatively greater compared to the variation within the cluster, so that the clusters formed are more compact and have minimal overlap.

Furthermore, for the Calinski-Harabasz Index which reaches the maximum value of K = 8, it shows that variation between clusters is much more dominant than variation within clusters. These findings confirm that the division of data into eight clusters is able to optimally describe the data structure. Taking into account the suitability and consistency of the results of the three validation indices, the optimal number of clusters in the data of subsidized fertilizer recipients in 2025 is set at eight clusters. The use of this multi-index evaluation approach provides a stronger and more comprehensive foundation compared to the determination of the number of clusters that rely on only one evaluation index.

2. Clustering Results Using K-Means++

After the optimal number of clusters is determined, *clustering analysis* is carried out using the K-Means++ algorithm with a number of clusters of eight. The evaluation of the quality of K-Means++ clustering results is shown in Table 2.

Table 2. Clustering Evaluation Results Using K-Means++

Evaluation Metrics	Value
Silhouette Score	0,756
Davies-Bouldin Index	0,241
Calinski-Harabasz Index	7901

The grouping results using the K-Means++ algorithm with an optimal number of clusters of eight showed excellent *clustering* quality , as presented in Table 2. A Silhouette Score value of 0.756 indicates that the separation between clusters is clearly formed and most of the fertilizer recipient data is in clusters that correspond to their respective characteristics.

A low value of the Davies-Bouldin Index indicates that the clusters formed have a high level of internal uniformity as well as noticeable differences between clusters. This condition reflects the ability of K-Means++ to produce compact and stable clusters. In addition, the high value of the Calinski-Harabasz Index further corroborates that the resulting cluster structure is at an optimal level of separation.

The performance advantages of K-Means++ are influenced by a better centroid initialization mechanism compared to conventional K-Means. The process of selecting a more dispersed initial cluster center allows K-Means++ to reduce the risk of unrepresentative or biased clusters forming. In the context of subsidized fertilizer recipients, these findings show that K-Means++ is able to group recipients based on differences in land area characteristics and fertilizer needs more clearly, so that *clustering results* are easier to interpret and relevant to support the formulation of subsidized fertilizer distribution policies.

3. Clustering Results Using Fuzzy C-Means

In addition to K-Means++, this study also applies Fuzzy C-Means (FCM) as a *soft clustering approach*. The results of the clustering quality evaluation using FCM are presented in Table 3.

Table 3. Clustering Evaluation Results Using Fuzzy C-Means

Evaluation Metrics	Value
Silhouette Score	0,737
Davies-Bouldin Index	0,424
Calinski-Harabasz Index	4903

The application of the Fuzzy C-Means (FCM) algorithm resulted in *a relatively good clustering* quality , although quantitatively it was still below the results obtained using K-Means++. The Silhouette Score value of 0.737 shows that the structure of the cluster formed is quite clear, but the level of separation between clusters is not as firm as the results of the K-Means++ grouping. This condition indicates the existence of a number of fertilizer recipient data that have characteristics adjacent to more than one cluster.

The higher value of the Davies-Bouldin Index than the K-Means++ indicates that the clusters generated by FCM have a comparatively greater degree of overlap. However, this condition also reflects the characteristics of the data of subsidized fertilizer recipients which are heterogeneous. In a real context, there are farmers with land area and fertilizer use patterns that are between two groups, so that *soft clustering* approaches such as FCM are able to represent these conditions through the value of membership degrees in each cluster.

Therefore, although numerically FCM results in lower *clustering* quality , it still provides an additional perspective in understanding data structures that are not completely separated by force. These findings suggest that FCM is still relevant for exploratory analysis, especially when the purpose of the study is to examine the proximity and overlap of characteristics between clusters.

4. Comparison of K-Means++ and Fuzzy C-Means

A comparison of the performance between the K-Means++ and Fuzzy C-Means algorithms was carried out to assess the most suitable method in the grouping of subsidized fertilizer recipients. A summary of the comparative performance of the two methods is presented in Table 4.

Table 4. Comparison of K-Means++ and Fuzzy C-Means Performance

Metric	K-Means++	Fuzzy C-Means
Silhouette Score	0,756	0,737
Davies-Bouldin Index	0.241	0,424
Calinski-Harabasz Index	7901	4903

The performance comparison between K-Means++ and Fuzzy C-Means shows a fairly clear difference, as summarized in Table 4. K-Means++ consistently produces higher Silhouette Score and Calinski-Harabasz Index values as well as lower Davies-Bouldin Index values than FCM. This shows that K-Means++ is superior in forming separate clusters clearly and stably.

In terms of policy implementation, *the* results of K-Means++ clustering are easier to interpret because each fertilizer recipient is only classified into one cluster. This condition makes it easier for stakeholders to design a more targeted and operational subsidized fertilizer distribution policy. In contrast, FCMs offer flexibility through membership degrees, but their interpretation is comparatively more complex for practical decision-making needs.

Overall, the results of this study show that K-Means++ is a more suitable method to be used in the grouping of subsidized fertilizer recipients in 2025, especially when the purpose of the analysis is to support policy decision-making that requires clear results and is easily translated into action.

Conclusion

This study was conducted to group the recipients of subsidized fertilizers in 2025 using the K-Means++ and Fuzzy C-Means methods, as well as determine the optimal number of clusters through a multi-index evaluation approach. The results showed that the most optimal number of clusters was eight clusters, which was obtained from the consistency of the values of the Silhouette Score, the Davies-Bouldin Index, and the Calinski-Harabasz Index.

The results of *the clustering* analysis showed that the K-Means++ method produced better grouping quality than the Fuzzy C-Means. The higher values of the Silhouette Score and the Calinski-Harabasz Index, accompanied by the lower values of the Davies-Bouldin Index, indicate that K-Means++ is able to form a clearer, more stable, and well-separated cluster. In addition, the results of the K-Means++ grouping are easier to understand because each fertilizer recipient is classified into one cluster, making it more practical to support the implementation of subsidized fertilizer distribution policies.

On the other hand, Fuzzy C-Means still contributes to understanding the characteristics of heterogeneous data through a *soft clustering approach*. This method is able to describe the proximity of characteristics between clusters through the degree of membership, so it is useful as an exploratory analysis tool for follow-up studies that require a deeper understanding of the data structure.

Overall, this study confirms that the application of K-Means++ with multi-index evaluation is an effective approach in the grouping of subsidized fertilizer recipients. The findings of this study are expected to be input for stakeholders in designing subsidized fertilizer distribution policies that are more targeted. Further research is recommended to consider the addition of variables and other analysis methods to improve the accuracy and breadth of grouping results.

References

Aldino, A. A., Darwis, D., Prastowo, A. T., & Sujana, C. (2021). Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency. *Journal of Physics: Conference Series*, *1751*(1). https://doi.org/10.1088/1742-6596/1751/1/012038

Daoudi, S., Zouaoui, C. M. A., El-Mezouar, M. C., & Taleb, N. (2021). Parallelization of the K-Means++ clustering algorithm. *Ingenierie Des Systemes d'Information*, *26*(1), 59–66. https://doi.org/10.18280/isi.260106

Hermantria, I. (2021). Segmentation of Non-Subsidized Fertilizer Consumers in East Nusa Tenggara Province. *BALANCE: Economic, Business, Management and Accounting Journal*, *18*(2), 8. https://doi.org/10.30651/blc.v18i2.7301

Ika Ari Sasmita, Rini Indriati, & M. Najibulloh Muzaki. (2021). Recommendations for Recipients of the Keluarga Harapan Program. *Jambura Journal of Electrical and Electronics Engineering*, *3*, 84–88.

Kodinariya, T. ., & Makwana, P. . (2021). Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, *1*(6), 90–95.

Krasnov, D., Davis, D., Malott, K., Chen, Y., Shi, X., & Wong, A. (2023). Fuzzy C-Means Clustering: A Review of Applications in Breast Cancer Detection. *Entropy*, *25*(7). https://doi.org/10.3390/e25071021

Kumar, S., & Singh, M. (2022). Collaborative annealing power K-means++ clustering. *Knowledge-Based Systems*.

Meng, J., Yu, Z., Cai, Y., & Wang, X. (2023). K-Means++ Clustering Algorithm in Categorization of Glass Cultural Relics. *Applied Sciences (Switzerland)*, *13*(8). https://doi.org/10.3390/app13084736

Prakash, L. N. C. K., Suryanarayana, G., Ansari, M. D., & Gunjan, V. K. (2022). Instantaneous approach for evaluating the initial centers in the agricultural databases using k-means clustering algorithm. *Journal of Mobile Multimedia*, *18*(1), 43–60. https://doi.org/10.13052/jmm1550-4646.1813

Putu Gede Subhaktiyasa, Sang Ayu Ketut Candrawati, N. Putri Sumaryani, Ni Wayan Sunita, & Abd. Syakur. (2025). Application of Descriptive Statistics: Quantitative and Qualitative Perspectives. *Emasains : Journal of Mathematics and Science Education*, *14*(1), 96–104. https://doi.org/10.59672/emasains.v14i1.4450

Seyed Emadedin Hashemi, Fatemeh Gholian-Jouybari, & Mostafa Hajiaghaei-Keshteli. (2023). A fuzzy C-means algorithm for optimizing data clustering. *Expert Systems with Applications*, *227*(120377).

Swain, K. P., Nayak, S. R., Ravi, V., Mishra, S., Alahmadi, T. J., Singh, P., & Diwakar, M. (2024). Empowering Crop Selection with Ensemble Learning and K-means Clustering: A Modern Agricultural Perspective. *The Open Agriculture Journal*, *18*(1). https://doi.org/10.2174/0118743315291367240207093403

Zalfa, D., Nugroho, A., & Harini, D. (2023). Clustering analysis in subsidized fertilizer recipients using the K-Means algorithm. In *August* (Vol. 7).

Zhang, Liu, Y., & Wang, J. (2023). Cluster Validation Techniques for Unsupervised Learning: A Comparative Study. *Expert Systems with Applications*.