Performance Comparison of Fuzzy C-Means and Decision Tree Algorithms for Analyzing Social Assistance Recipient Eligibility

Riska Wirdayanti [1]*, Dahlan Abdullah [2]*, Nurdin [3]*
* Department of Information Technology, Universitas Malikussaleh, Lhokseumawe, Indonesia
riska.227110201015@mhs.unimal.ac.id [1], dahlan@unimal.ac.id [2], nurdin@unimal.ac.id [3]

Article Info

Article history:

Received 2026-02-05
Revised 2026-03-07
Accepted 2026-04-10

Keyword:

Fuzzy C-Means,
Decision Tree,
Social Assistance,
PKH,
Classification.

ABSTRACT

The Family Hope Program (PKH) is one of the government's social assistance programs aimed at improving the welfare of low-income communities. In the determination process, an accurate and measurable analysis method is required so that the selection of potential beneficiaries can be carried out objectively. This study aims to compare the performance of the Fuzzy C-Means (FCM) method and the Decision Tree algorithm in analyzing the eligibility of PKH social assistance recipients. The FCM method is used to group data unsupervised into eligible and ineligible clusters, while the Decision Tree is used as a supervised classification method by utilizing socioeconomic attributes as input variables. The testing was conducted using accuracy, precision, recall, F1-Score, and computation time metrics. The dataset uses from the Targeting for the Acceleration of Extreme Poverty Eradication (P3KE) program in Binjee Village, Nisam District, North Aceh Regency, with a total sample size of 632 individual records. The results showed that FCM was able to form two clusters with an accuracy value of 59.95% and an F1-Score of 70.03%. Meanwhile, Decision Tree performed much better with an accuracy of 94.96%, an F1-Score of 95.04%, a precision of 92.33%, a recall of 97.41%, and a computation time of 0.0664 seconds. Based on these evaluation results, it can be concluded that Decision Tree is more effective than FCM because it is capable of producing higher accuracy, more consistent evaluation metric stability, clear classification rule interpretation, and comparable processing time efficiency. Thus, the Decision Tree algorithm is considered more optimal and suitable for use in supporting the process of analyzing the eligibility of PKH social assistance recipients in an objective and measurable manner.

I. INTRODUCTION

The Family Hope Program (PKH) is a conditional social assistance program provided to poor and vulnerable families registered in the Integrated Social Welfare Data (DTKS). This program aims to enhance the quality of human resources by providing improved access to healthcare and education services. However, the implementation of this program faces various challenges, particularly in ensuring that assistance is distributed to those who are truly eligible [1][2]. Targeting inaccuracies can arise from a lack of recipient data validation, selection processes that are still conducted manually, and limitations in the use of supporting technology. Such conditions not only hinder program effectiveness but also have the potential to trigger public dissatisfaction with government policies. Inaccuracy in determining social assistance recipients can diminish public trust in the ongoing programs. Furthermore, the misallocation of funds can lead to budget waste and reduce overall program efficiency [3].

In the context of social assistance, data mining methods can be employed to enhance accuracy in selecting recipients by analyzing recipient data based on relevant characteristics. Data Mining is an analytical process used to extract information from large datasets to identify hidden patterns and present them in a format useful for decision-making [4] [5]. One of the clustering methods that can be utilized is Fuzzy

C-Means (FCM), which is capable of grouping data based on membership levels and is effective in handling complex data [6][7].

In addition, the Decision Tree classification method offers the advantage of producing transparent and comprehensible models, making it suitable for decision-making contexts involving multiple stakeholders, while providing a clear and easy-to-understand classification structure [8][9]. The Decision Tree is also characterized by its interpretability and efficiency in the classification process. It is capable of delivering satisfactory results through a model structure in the form of a decision tree that visualizes the classification process, thereby facilitating easier analysis and implementation [10][11].

This study aims to compare the performance of the Fuzzy C-Means method and the Decision Tree algorithm in the selection process of social assistance recipients. Furthermore, this research seeks to support the primary objectives of social assistance programs improving public welfare through a more measurable and data-driven approach. The analytical results obtained from this study will not only provide recommendations regarding the most effective method but also serve as a foundation for developing data-driven policies that are more responsive to public needs. This is in line with the findings of Ahmat Arifin, who emphasizes the importance of utilizing data analysis technology to enhance the accuracy and efficiency of social assistance programs [12][13].

II. METHODOLOGY

A. Research Stages

In this study, a series of systematic stages are employed to ensure the smooth execution and success of the research, as well as to obtain valid results that align with the research objectives.

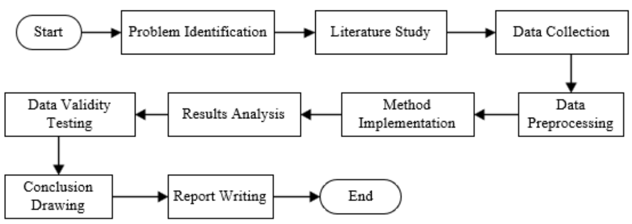

Figure 1. Research Framework

The following is a detailed explanation of the stages conducted in this research:

1. Problem Identification: Identifying issues with the selection process for PKH social assistance recipients and defining the research objectives.
2. Literature Study: Conducting a review of relevant previous studies concerning the Fuzzy C-Means (FCM) method and the Decision Tree algorithm in data analysis.
3. Data Collection: Gathering demographic and socio-economic data from residents in Binjee Village, Nisam District, North Aceh Regency.
4. Data Preprocessing: Performing data cleaning, normalization, and transformation to ensure the quality and readiness of the data for analysis.
5. Method Implementation: Applying the Fuzzy C-Means (FCM) method for clustering and the Decision Tree algorithm for the classification of social assistance recipients.
6. Results Analysis: Evaluating the outcomes of both methods based on Confusion Matrix.
7. Data Validity Testing: Testing the analytical results to ensure the effectiveness of the methods used in determining the eligibility of social assistance recipients.
8. Conclusion Drawing: Formulating conclusions based on the analytical results and comparing the strengths and weaknesses of each method.
9. Research Report Writing: Documenting the entire research process, findings, and conclusions into a Research Paper.

B. Fuzzy C-Means (FCM)

Fuzzy C-Means (FCM) is a data mining clustering method with the unsupervised learning category that allows a single data point to belong to more than one cluster based on a specific degree of membership [14]. Each data point is assigned a membership value, indicating the extent to which that point fits into each respective cluster. This approach is highly suitable for complex data or datasets that lack clear, rigid boundaries between specific groups.

The Iterative Process of the Fuzzy C-Means AlgorithmThe implementation of the Fuzzy C-Means (FCM) algorithm follows a systematic iterative procedure to partition the dataset into optimal clusters [15]. The process begins with the initialization of membership degrees (u_ij). where each data point is assigned an initial value representing its association with the clusters. Following this initialization, the algorithm enters an iterative loop.The first step within the iteration is to update the cluster centers (c_j). These centroids are recalculated based on the weighted average of all data points, where the influence of each point is determined by its membership degree. Subsequently, the algorithm updates the membership degrees (u_ij) for each data point by evaluating its relative distance to the newly calculated centroids.This cycle of updating centroids and membership values is repeated until convergence is achieved. Convergence occurs when the change in the objective function between iterations falls below a predefined tolerance threshold or when the maximum number of iterations is reached, signifying that the cluster centers have stabilized and an optimal partition has been found [16]. The mathematical formulation underpinning the Fuzzy C-Means algorithm is defined as follows, primarily focusing on the minimization of the objective function [17] [18].

1. Objective Function (J_m) aims to minimize the distance between data points and cluster centers, multiplied by the membership weight of each data point:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \, ||x_i - c_j||^2 \qquad (1)$$

Description:

J_m : The objective function to be minimized.

N : Total number of data points.

C : Number of clusters.

x_i : The i-th data point.

c_j : The center of the j-th cluster (centroid).

u_ij : Degree of membership of data x_i in cluster j.

m : Fuzziness parameter (typically set to m > 1) to control the level of fuzziness.

‖x_i-c_j ‖2: The squared Euclidean distance between data point x_i nd cluster center c_j.

2. The Membership Degree Update (Uij) quantifies the probability of a data point's association with a specific cluster, determined by its relative distance to all existing cluster centroids.

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{||x_i - c_j||}{||x_i - c_k||} \right)^{\frac{2}{m-1}}} \qquad (2)$$

3. Update Pusat Cluster (Cj) is computed as the weighted average of all data points assigned to the cluster, incorporating the membership degrees as weighting factors.

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m} \qquad (3)$$

C. Decision Tree

The Decision Tree algorithm is a classification method employed to facilitate decision-making through the analysis of available datasets. This algorithm constructs a tree-like structure that illustrates a sequence of decisions based on specific attributes, such as income, number of dependents, age, and other relevant variables. The tree structure consists of nodes, where each node represents a decision-making point based on a particular condition or attribute [19]. The illustration of the decision tree is presented below.

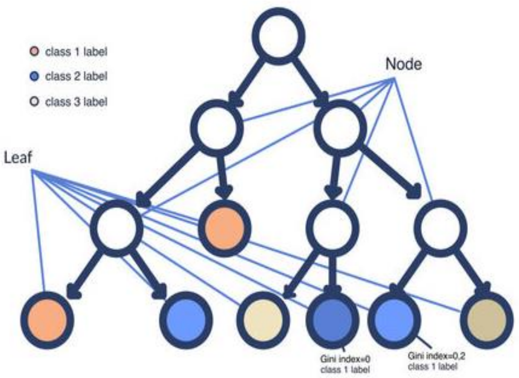

Figure 2. Decision Tree Structure

The fundamental concepts of the Decision Tree algorithm are as follows [20]:

1. Impurity: The Decision Tree partitions the dataset based on attributes that minimize impurity. This measure determines how mixed the classes are within a specific node. To quantify this impurity, two commonly utilized metrics are employed.

2. Entropy: Entropy is used to measure the degree of disorder or impurity within a particular group of data. In the context of classification, it quantifies the uncertainty of the information, where a lower entropy value indicates a more homogenous or "pure" group.

$$Entropy = -\sum_{i=1}^{k} p_i \, log_2(p_i) \qquad (4)$$

Description:

p_i: The proportion of data points belonging to the *i*-th class.

k: The total number of classes.

3. Gini Index (CART) The Gini Index serves as an alternative metric for measuring impurity by calculating the probability of a specific element being incorrectly classified if it were randomly labeled according to the distribution of labels in the subset. It is the primary metric used in the Classification and Regression Tree (CART) algorithm.

$$Gini = 1 - \sum_{i=1}^{k} p_i^2 \qquad (5)$$

4. Information Gain (IG) Information Gain quantifies the reduction in uncertainty (impurity) achieved after a dataset is partitioned. It measures the decrease in entropy following the split based on a specific attribute. In the tree construction process, the attribute that yields the highest Information Gain is selected as the splitting node.

$$IG = Entropy(Parent - \sum_{j=1}^{k} \frac{N_j}{N} Entropy(Child_j) \qquad (6)$$

Description:

N_j: The number of data points in branch

N: The total number of data points (total observations).

5. Splitting Rule The data partition is executed based on the attribute that yields the optimal Information Gain or the lowest Gini Index, effectively minimizing the impurity of the resulting subsets.

Decision Tree Algorithm first steps are select the optimal attribute to Identify the attribute with the highest Information Gain or the lowest Gini Index, Construct the decision node to Establish a node based on the selected attribute, Partition the dataset: Split the data into branches corresponding to the specific values of the attribute, at last the Iterate process do repeat these steps until all data points are classified or the stopping criteria are met (e.g., impurity approaches zero or the tree reaches a predefined maximum depth) [21].

D. Confusion Matrix

The Confusion Matrix is a fundamental evaluation metric used to assess the performance of classification models in machine learning. It serves as a straightforward yet effective

method for evaluating the predictive power of a system [22]. This technique aims to measure the performance level or accuracy of the system when classifying testing data. The matrix provides a detailed summary of the number of correct and incorrect predictions made by the model, comparing the predicted outcomes against the actual labels. The Confusion Matrix consists of four primary components:
1. True Positive (TP): The number of instances that are actually positive (Eligible) and are correctly predicted as positive.
2. False Positive (FP): The number of instances that are actually negative (Not Eligible) but are incorrectly predicted as positive (Eligible).
3. True Negative (TN): The number of instances that are actually negative (Not Eligible) and are correctly predicted as negative.
4. False Negative (FN): The number of instances that are actually positive (Eligible) but are incorrectly predicted as negative (Not Eligible).

Based on these four values, various evaluation metrics—including Accuracy, Precision, Recall, and F1-Score—can be calculated to provide a more objective and comprehensive assessment of the model's performance.

III. RESULT AND DISCUSSION

A. Data Preprocessing Stage

In this study uses a dataset from the Targeting for the Acceleration of Extreme Poverty Eradication (P3KE) program in Binjee Village, Nisam District, North Aceh Regency, with a total sample size of 632 individual records. The dataset is sensitive and confidential. Consequently, some original data has been redacted, and the research results are limited to quantitative values to protect the privacy of the individuals involved. The variables in this research represent the criteria for potential social assistance recipient families. The eight variables used are:

- Gender
- Relationship to the head of the household
- Welfare decile
- Marital status
- Occupation type
- Employment status
- Educational background
- Age

Prior to the classification and clustering processes, the raw data underwent a preprocessing stage to ensure it was suitable for analysis. The objective of preprocessing is to clean, organize, and transform the data into a numerical format compatible with the algorithms. The preprocessing steps implemented in this study are detailed as follows:
1. Feature Selection (Column Filtering) From an initial set of 32 columns, 10 relevant feature columns were selected based on their significance to the PKH (The Family Hope Program) social assistance eligibility analysis. Following this process, the dataset dimensions were refined to 632 rows × 10 columns, resulting in a total of 6,320 data cells.
2. Missing Value Analysis The analysis identified 1,400 missing cells out of the total 6,320, representing a missing data percentage of 18.13%. These missing values were handled using specific strategies tailored to the column types, such as default value imputation or specific encoding for missing categories.
3. Categorical Data Encoding This process was conducted to convert categorical values into numerical representations, enabling the data to be processed by the algorithms.

TABLE I.
CATEGORICAL COLUMN ENCODING

No.	Column Name	Unique Categories	Category Mapping	Encoding Range
1	Gender	2	{'Male': 0, 'Female': 1, NaN: 2}	**0 – 2**
2	Relationship to Head of Household	4	{'Child': 0, 'Spouse': 1, 'Head of Household': 2, 'Others': 3, NaN: 4}	**0 – 4**
3	Marital Status	2	{'Single': 0, 'Married': 1, NaN: 2}	**0 – 2**
4	Occupation	7	{'Farmer': 4, 'Unemployed': 5, 'Entrepreneur': 6, 'Private Employee': 2, 'Fisherman': 0, ...}	
5	Employment Status	1	{'Empty': 0, NaN: 1}	**0 – 1**
6	Education Level	10	{'No Schooling': 0, 'Elementary School': 1, 'Junior High School': 2, 'Senior High School': 3, ...}	**0 – 10**

B. Implementation of Fuzzy C-Means (FCM)

The Fuzzy C-Means (FCM) implementation was performed on the preprocessed real-world dataset with dimensions of 1186 × 8. In this stage, the 'Age' column was standardized using Z-score normalization, resulting in a transformation range from [-2.1369 to 3.2235]. The initial parameters configured for the FCM algorithm are presented below.

TABLE II.
INITIAL FCM PARAMETERS

Parameter	Value
Number of clusters (c)	2
Fuzziness exponent (m)	2
Total number of data (N)	1186
Number of features (n)	8

Following the membership initialization process, all attributes were calculated to determine the cluster center values for C1 and C2.

TABLE III.
FINAL CLUSTER CENTERS

Attribute	**C1 (*Cluster* 0)**	**C2 (*Cluster* 1)**
Gender	0.7843	0.5268
Relationship to Household Head	0.9583	0.3468
Marital Status	0.8319	0.2746
Occupation	4.8661	4.8400
Employment Status	0.0000	0.0000
Education Level	6.4424	2.5234
Age	0.2000	-0.8023
Welfare Decile	1.6055	1.5466
Year	2021.0000	2021.0000

Furthermore, in the application of Fuzzy C-Means to the real-world dataset, the Euclidean distance between centroids was calculated to assess the degree of separation between the two formed clusters. Based on the calculation, a distance value of 4.1374 was obtained. This distance indicates that the two centroids exhibit a significant difference within the multidimensional attribute space. Consequently, it can be concluded that the generated clusters possess a high degree of separation and represent distinct data characteristics.

To ensure that the Fuzzy C-Means iterative process has reached a stable result, a convergence check was performed on the cluster center values. Convergence occurs when the change in cluster centers in subsequent iterations becomes negligible, signifying that no further significant adjustments are required. The algorithm terminates when the change in cluster centers is less than a predefined threshold (typically 0.001). The observed changes in cluster centers are as follows:

a. Gender C1: |0.51 - 0.50| = 0.01
b. Relationship to Head of Household C1: |0.47 - 0.43| = 0.04.
c. And so on...

If the change remains greater than the threshold, the process continues to the next iteration.

TABLE IV.
CONVERGENCE CHECK

Atribut	**C1**	**C2**
Gender	0.25	1
Relationship to Household Head	0.625	0.125
Marital Status	0.75	0.25
Occupation	0	1
Education Level	0.46875	0.71875
Age	0.70625	0.3375
Welfare Decile	0.166666667	0.333333

The iterative process terminates once convergence is achieved.

TABLE V.
MAPPING CLUSTERS TO ACTUAL LABELS

Cluster	**Total Data**	**Label 1 (Eligible)**	**Label 0 (Not Eligible)**	**Dominant Label**	**Dominance Percentage**
1	992	555	437	1 (Eligible)	55.95%
2	194	38	156	0 (Not Eligible)	80.41%

This study was conducted through ten independent trials to compare the performance of the conventional K-Medoids algorithm with the optimized version employing Z-Score-based initialization. In each trial, the initial medoids were selected from the three data points with the lowest Z-Score values. These instances were assumed to lie closest to the center of the distribution and, therefore, serve as effective initial representatives for the clusters. The selected medoids for each trial are summarized in the following table.

Table 2 presents the results of the first trial using the Z-Score-based initialization method in the K-Medoids clustering algorithm. The initial medoids (data points 79, 679, and 756) were selected based on the lowest Z-Score values. In the first iteration, the medoids were updated to a new combination (725, 496, and 767), resulting in a reduced total minimum distance, indicating an improvement in clustering compactness. However, in the second iteration, the total distance increased again, suggesting a potential shift away from the optimal configuration. This iterative behavior reflects the dynamic adjustment of medoids based on intra-cluster similarity, and such changes are evaluated across multiple trials to assess the overall stability and effectiveness of the proposed approach.

TABLE II.
RESULTS OF TRIAL 1 USING Z-SCORE-BASED INITIALIZATION

Medoid Selection Step	**Data Point Numbers**	**Total Minimum Distance**	**Distance Difference**
Initial Medoids	79, 679, 756	2,102,220.48	–
Iteration 1	725, 496, 767	1,954,349.92	-147,870.56
Iteration 2	784, 674, 548	2,082,161.21	127,811.29

C. *Decision Tree Implementation*

The Decision Tree algorithm is a supervised learning method that operates by constructing a decision tree based on the most relevant information extracted from the input data. In this study, the Decision Tree is utilized to build a classification model for social assistance eligibility based on the preprocessed dataset.

Before constructing the decision tree, an analysis was performed on the initial data distribution based on the labels (Recipient = 1 and Non-Recipient = 0). The Information Gain for various features was calculated to determine the best attribute for data splitting, as shown in the following table.

TABLE VI.
INFORMATION GAIN RESULTS FOR EACH ATTRIBUTE

No	Feature	Parent Entropy	Weighted Entropy	Information Gain
1	Gender	0.9999	0.6332	0.3667
2	Relationship to Head of Household	0.9999	0.3607	0.6392
3	Marital Status	0.9999	0.6146	0.3853

The feature 'Relationship to Head of Household' possesses the highest Information Gain (0.6392) and serves as the potential root node for the decision tree. Subsequently, an Information Gain calculation was performed on the six tested features, namely Gender, Relationship to Head of Household, Marital Status, Occupation, Welfare Decile (Desil), and Age.

TABLE VII.
RESULTS OF ENTROPY, INFORMATION GAIN, AND NODE DECISIONS

Feature (Attribute)	Attribute Entropy	Information Gain	Decision / Rule
Gender	0.6068	0.2044	Level 1: Relationship = 0 → Non-Recipient
Relationship to Head of Household	0	0.8113	Level 1: Relationship = 1 → Recipient
Marital Status	0.0613	0.75	Level 2: Relationship = 2, Decile = 1 → Non-Recipient
Occupation	0	0.8113	Level 2: Relationship = 2, Decile = 2 → Recipient
Welfare Decile (Desil)	0	0.8113	-
Age	0.5	0.3113	-

Based on these results, the "Relationship to Head of Household" feature once again exhibits the highest Information Gain (0.8113). This finding is consistent with the results obtained during the processing of the entire dataset, reinforcing its position as the primary candidate for the root node. Furthermore, the Occupation and Welfare Decile features provide maximum Information Gain at subsequent levels when data splitting is performed for the "Relationship = 2" category. This further supports the final classification decisions for "Recipient" and "Non-Recipient" status. The decision tree structure derived from these manual calculations generally demonstrates the data partitioning flow as follows:

a. If Relationship = 1, then the data is predicted as Recipient.

b. If Relationship = 0, then the data is predicted as Non-Recipient.

c. If Relationship = 2, then a further split is performed based on the Welfare Decile:

• Decile = 1 → Non-Recipient

• Decile = 2 → Recipient

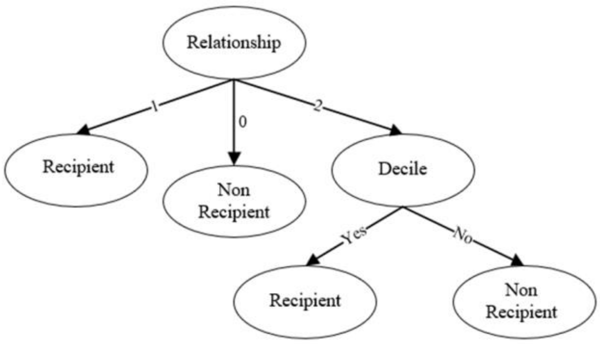

Figure 2. Decision Tree

D. *Evaluation and Confusion Matrix of the Fuzzy C-Means*

To evaluate the clustering results obtained using the Fuzzy C-Means (FCM) method, a mapping of the generated clusters to the actual labels was performed based on the label distribution within each cluster. The objective is to assess the extent to which the clustering results represent the original classes or ground truth labels.

TABLE VIII.
CONFUSION MATRIX FUZZY C-MEANS

	Predicted: Not Eligible	Predicted: Eligible
Actual: Not Eligible	(TN) 156	(FP) 437
Actual: Eligible	(FN) 38	(TP) 555

This confusion matrix serves as the basis for calculating model performance metrics, including accuracy, precision, recall, specificity, and F1-score, to assess how effectively the FCM method distinguishes between Eligible and Not Eligible data. Based on this distribution, it is observed that the FCM method performs reasonably well in identifying Eligible data (as indicated by the high TP value). However, it still produces a significant number of misclassifications for the Not Eligible data (indicated by the high FP value), which subsequently results in lower precision and specificity scores.

TABLE VIII.
CONFUSION MATRIX FUZZY C-MEANS

Metric	Formula	Value
Accuracy	(TP + TN) / (TP + TN + FP + FN)	0.5995 (59.95%)
Precision	TP / (TP + FP)	0.5595 (55.95%)
Recall (Sensitivity)	TP / (TP + FN)	0.9359 (93.59%)
Specificity	TN / (TN + FP)	0.2631 (26.31%)
F1-Score	2 × (*Precision* × *Recall*) / (*Precision* + *Recall*)	0.7003 (70.03%)

The evaluation results indicate that the Fuzzy C-Means (FCM) method demonstrates strong performance in terms of Recall, achieving a value of 93.59%. This suggests that the majority of data points belonging to the 'Eligible' category were successfully identified by the model. In other words, this method is highly sensitive in detecting 'Eligible' cases. However, the Precision value of only 55.95% and the low Specificity of 26.31% reveal that the model tends to generate a high number of False Positives—where data points that are actually 'Not Eligible' are predicted as 'Eligible'. This indicates that FCM tends to overpredict the 'Eligible' class, assigning significantly more data to this category than is reflected in the actual ground truth.

E. Evaluation and Confusion Matrix of the Decision Tree

After the Decision Tree model training process was completed, an evaluation of the classification performance was conducted using a testing dataset of 238 records. The evaluation results are presented in the form of a confusion matrix and performance metrics.

TABLE IX.
CONFUSION MATRIX DECISION TREE

	Predicted: Not Eligible	Predicted: Eligible
Actual: Not Eligible	True Negative (TN) = 111	False Positive (FP) = 12
Actual: Eligible	False Negative (FN) = 0	True Positive (TP) = 115

TABLE X.
CLASSIFICATION PERFORMANCE METRICS FOR DECISION

Evaluation Metric	Formula	Value
Accuracy	(TP + TN) / (TP + TN + FP + FN)	0.9496 (94.96%)
Precision	TP / (TP + FP)	0.9055 (90.55%)
Recall (Sensitivity)	TP / (TP + FN)	1.0000 (100%)
Specificity	TN / (TN + FP)	0.9024 (90.24%)
F1-Score	2 * (Precision * Recall) / (Precision + Recall)	0.9504 (95.04%)

The model demonstrates excellent performance, characterized by high accuracy (94.96%), a strong F1-score (95.04%), and zero errors in classifying actual aid recipients (FN = 0). This indicates that the Decision Tree is highly effective in handling classification tasks for data that has been previously clustered.

F. Comparison of FCM and Decision Tree

After evaluating both methods, namely Fuzzy C-Means (FCM) and Decision Tree, the performance metric results can be compared across several key aspects: accuracy, precision, sensitivity, F1-score, computational time, model complexity, and interpretability.

TABLE XI.
COMPARISON OF EVALUATION METRICS

Evaluation Metric	Decision Tree	Fuzzy C-Means	Difference
Accuracy	0.9496 (94.96%)	0.5995 (59.95%)	0.3501
Precision	0.9055 (90.55%)	0.5595 (55.95%)	0.3460
Recall	1.0000 (100.0%)	0.9359 (93.59%)	0.0641
F1-Score	0.9504 (95.04%)	0.7003 (70.03%)	0.2501

The Decision Tree method outperforms Fuzzy C-Means across all performance evaluation metrics, particularly in precision and specificity, which signifies the Decision Tree's ability to minimize misclassifications within the 'not eligible' group. The evaluation results from the confusion matrix can be visualized as follows:

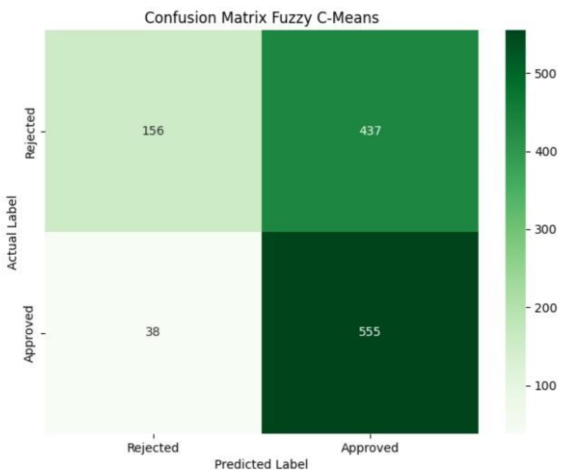

Figure 2. Confusion Matrix Fuzzy C-Means

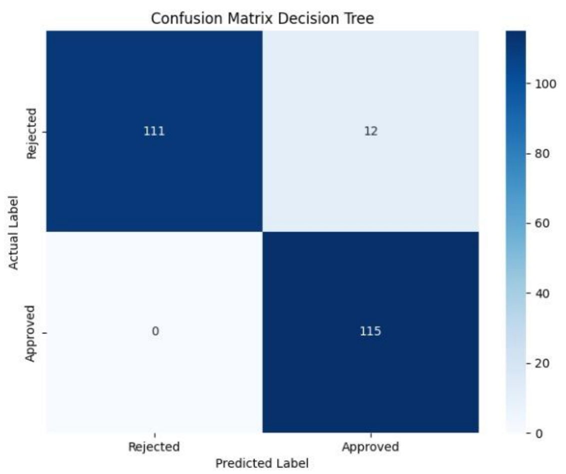

Figure 3. Confusion Matrix Decision Tree

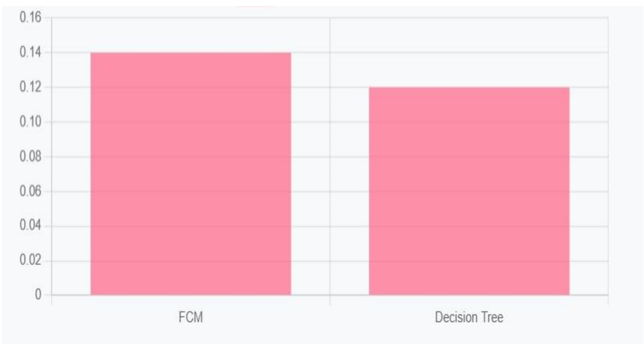

Figure 4. Processing Speed

Processing speed is calculated based on the number of samples processed per second. Measurement results indicate that the Decision Tree algorithm requires 0.1230 seconds, accounting for approximately 47.08% of the total computational time. Meanwhile, the Fuzzy C-Means method requires a slightly longer duration of 0.1382 seconds, or 52.92% of the overall time. Other preprocessing stages do not demand significant additional computational time, recorded at 0 seconds (0%). Consequently, the total computational time required for the entire analytical process is 0.2612 seconds.

IV. CONCLUSION

The performance of the Fuzzy C-Means method demonstrates its capability to group data into two clusters: "eligible" and "not eligible" for social assistance. However, the achieved accuracy of 59.95% and an F1-score of 70.03% indicate that this method is not yet optimal for use as a classification baseline in the context of datasets that possess final decision labels. The performance of the Decision Tree algorithm shows excellent results, with accuracy reaching 94.96% and an F1-score of 95.04%. Furthermore, the Decision Tree provides a clear interpretation of the generated classification rules, making it highly suitable for decision-making processes involving socio-economic criteria.

Based on the evaluation and comparison results, it can be concluded that the Decision Tree algorithm demonstrates more effective performance compared to the Fuzzy C-Means method in analyzing the eligibility of Family Hope Program (PKH) social assistance recipients. This is evidenced by the Decision Tree's accuracy of 94.96%, an F1-Score of 95.04%, precision of 92.33%, recall of 97.41%, and a computational time of 0.0664 seconds. Conversely, the Fuzzy C-Means method only obtained an accuracy of 59.95%, an F1-Score of 70.03%, precision of 55.95%, recall of 93.59%, and a computational time of 0.0489 seconds. These differences in evaluation values indicate that the Decision Tree has far superior performance, provides clearer interpretation of classification rules, and remains efficient in processing, making it more appropriate for supporting the eligibility analysis of social assistance recipients.

Interpretability or the ease of explaining model results, is a crucial aspect of social research, particularly when the outcomes serve as a basis for public decision-making. The results indicate that the Decision Tree model has a clear advantage in interpretability, as it is based on explicit rules and can be visualized as a decision tree. Conversely, the FCM (Fuzzy C-Means) model requires an advanced understanding of fuzzy membership logic and cluster centroids, which tends to be difficult for non-technical stakeholders to comprehend.

REFERENCES

[1] N. S. A. Aeni, G. Oktariyana, A. L. Puspita, A. Naila, and Y. A. Awdovi, "Implementasi Program Keluarga Harapan (PKH) di Kota Serang," *J. Ilmu Sos. Polit. dan Hum.*, vol. 7, no. 1, pp. 30–39, 2024, doi: 10.36624/jisora.v7i1.150.

[2] A. Sapitri, N. Nurdin, and Y. Afrilia, "Implementation of Clustering Method Using K-Means Algorithm for Grouping BPJS

Health Patient Medical Record Data," *J. Appl. Informatics Comput.*, vol. 9, no. 5, pp. 2391–2398, 2025, doi: 10.30871/jaic.v9i5.10046.
[3] S. A. Ajun, S. Canon, B. R. Payu, and U. N. Gorontalo, "Analisis ketepatan sasaran penerima bantuan sosial provinsi gorontalo," vol. 7, pp. 8–19, 2024, doi: 10.37600/ekbi.v7i1.1319.
[4] D. Amelia, D. Abdullah, and Z. Fitri, "Commodity Distribution In Central Aceh Through An Integrated Auction Market System Using The Android-Based Association Rule Mining (ARM) Method," vol. 2, pp. 2–6, 2024.
[5] N. Nurdin, "Analisa Data Mining Dalam Memprediksi Masyarakat Kurang Mampu Menggunakan Metode K-Nearest Neighbor," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 1090–1098, 2024, doi: 10.23960/jitet.v12i2.4131.
[6] N. Nurdin, S. Fitriani, Z. Yunizar, and B. Bustami, "Clustering the Distribution of COVID-19 in Aceh Province Using the Fuzzy C-Means Algorithm," *JTAM (Jurnal Teor. dan Apl. Mat.*, vol. 6, no. 3, p. 665, 2022, doi: 10.31764/jtam.v6i3.8576.
[7] W. An-naziz Safaat, R. Kurniawan, and Y. Arie Wijaya, "Penerapan Data Mining Clustering Menggunakan Algoritma Fuzzy C-Means Pada Data Pencari Kerja Di Kabupaten Kuningan," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1507–1511, 2024, doi: 10.36040/jati.v8i2.8411.
[8] R. D. Mandasari, B. S. Informatika, B. S. Informatika, D. B. Sosial, K. Kelayakan, and P. Keputusan, "Implementasi Algoritma Decision Tree Untuk," vol. 6, no. 2, pp. 240–246, 2024.
[9] W. R. Amalia, N. Nurdin, and Y. Afrillia, "Klasifikasi Penerima Bantuan Sektor Kelautan dan Perikanan dengan Metode Support Vector Machine," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 1919.
[10] M. K. Faruq, K. Umam, M. I. Mustofa, and A. A. Mahfudh, "Comparative Study of SVM and Decision Tree Algorithms on the Effect of SMOTE Technique on LinkAja Application," vol. 9, no. 6, pp. 3305–3311, 2025.
[11] M. Rahma, "Sistem Pendukung Keputusan Penentuan Penerimaan Bantuan PKH Menggunakan Metode Naïve Bayes," *TECHSI - J. Tek. Inform.*, vol. 14, no. 1, p. 21, 2023, doi: 10.29103/techsi.v14i1.6960.
[12] A. Arifin, F. Al Zami, and E. Norsasangko, "Prediksi P Enerima B Easiswa U Ang K Uliah T Unggal," vol. 2, pp. 45–52, 2024.
[13] N. Imanda and N. Nurdin, "Penerapan Algorima Naive Bayes Pada klasifikasi penduduk kurang mampu dan mampu di tanoh anou idi rayeuk," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024
[14] R. D. Putri and D. P. Sari, "Regional Clustering in Sumatera Based on Walfare Indicators Using Fuzzy C-Means," vol. 9, no. 5, pp. 2399–2409, 2025.
[15] D. S. Riana, F. Rakhmawati, and R. Widyasari, "Cluster Analysis to Classify The Level of Social Welfare of The Community in Deli Serdang Regency Using Fuzzy C-Mean," *J. Math. Sci. Comput. with Appl. Publ.*, vol. 2, no. 2, pp. 116–124, 2021.
[16] S. Huang and F. Zeng, "An intelligent decision support system for agriculture economy based on fuzzy c-means," *Int. Food Agribus. Manag. Rev.*, no. March, pp. 1–14, 2024, doi: 10.22434/ifamr1029.
[17] A. C. Putra and K. D. Hartomo, "Optimalisasi Penyaluran Bantuan Pemerintah Untuk UMKM Menggunakan Metode Fuzzy C-Means," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 10, pp. 474–482, 2021, [Online]. Available:https://jurnal.iaii.or.id/index.php/RESTI/article/view/2980/423
[18] M. Ridha, N. Nurdin, M. Maryana, "Clustering Analysis to Identify Stunting Vulnerability Areas in North Aceh District Using the Fuzzy C-Means Algorithm ," *J. Informatics Telecommun. Eng.*, vol. 7, no. July, pp. 71–83, 2025, doi: 10.31289/jite.v9i1.14892.
[19] N. Azwanti and N. E. Putria, "Analisis Kepuasan Customer pada Sdtechnology Computer dengan Algoritma Decision Tree," *J. Desain Dan Anal. Teknol.*, vol. 3, no. 2, pp. 137–148, 2024, doi: 10.58520/jddat.v3i2.62.
[20] D. Septhya *et al.*, "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 15–19, 2023, doi: 10.57152/malcom.v3i1.591.
[21] R. N. Ramadhon, A. Ogi, and A. P. Agung, "Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank," vol. 3, pp. 1860–1874, 2024.
[22] P. Algoritma, K. N. Dan, F. Ekstraksi, and N. D. Analisis, "KOMPUTA : Jurnal Ilmiah Komputer dan Informatika KOMPUTA : Jurnal Ilmiah Komputer dan Informatika," vol. 12, no. 1, pp. 91–100, 2023.