Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.2, April 2026, pp. 1446~1453
e-ISSN: 2548-6861

Comparison of K-Means and Fuzzy C-Means for Regional Clustering Based on Unmet Need Determinants in East Java

Arina Nihayata Husna [1*], Nurissaidah Ulinuha [2]**
* Mathematics, Science and Technology, UIN Sunan Ampel Surabaya, Indonesia
arinnihayaa00@gmail.com [1], nuris.ulinnuha@uinsa.ac.id [2]

Article Info

Article history:

Received 2026-02-03
Revised 2026-03-12
Accepted 2026-04-10

Keyword:

Clustering,
Fuzzy C-Means,
K-Means,
Unmet need.

ABSTRACT

Unmet need for family planning (KB) refers to a condition in which fertile couples wish to delay or stop pregnancy but are not using any contraceptive methods. In 2022, East Java recorded a relatively high level of unmet need, indicating the need for a data-driven strategy to improve the effectiveness of the family planning program. This study compares the performance of the K-Means and Fuzzy C-Means (FCM) algorithms in clustering regencies/cities based on key determinants of unmet need, namely the number of fertile-age couples, poverty index, median age at first marriage, and female labor force participation rate. Secondary data obtained from BKKBN in 2023 were processed using data normalization, clustering, and evaluation with the Silhouette Coefficient and Davies–Bouldin Index, followed by denormalization to interpret the clustering results. The results indicate that K-Means outperforms Fuzzy C-Means in clustering regions based on the analyzed characteristics. After four iterations, the clustering results classify 22 regencies/cities into a low unmet-need cluster, 3 regencies into a moderate unmet-need cluster, and 13 regencies/cities into a high unmet-need cluster, achieving a Silhouette Coefficient value of 0.6997 and a Davies–Bouldin Index of 0.3539.

I. INTRODUCTION

Family Planning (KB) is one of the Indonesian government's key programs aimed at improving the quality of life of families through encouraging fertile couples to employ contraception properly. However, there is a high unmet need for the KB problem wherein fertile couples desire to space or stop childbearing but are not practicing any form of contraception. This indicates a gap between the need for and use of contraception in society [1]. Government Regulation No. 52 of 2009 on Government Regulation on Population Development and Family Development further supports the implementation of family planning programs in Indonesia [2]. The large number of poor-quality citizens and population increase pressure on the government to emphasize the function of family planning programs as a way of achieving balance in sustainable development [3]. This notwithstanding, implementation of the Family Planning Program still faces challenges, one of which is the unmet need phenomenon in society [4].

Unmet need for family planning (KB) is a condition in which couples who are fertile desire to delay or avoid pregnancy but do not use contraceptive methods [5]. Lack of use of contraceptives could be due to poor access, fear of side effects, spouse disapproval, religiosity, or unmet women's reproductive rights [6]. This state reflects an unaddressed necessity in the KB program, which, if not seriously addressed, can have impacts on household well-being, such as higher maternal mortality risk and stunting of children, higher costs to families' economic resources, a high TFR that can erode demographic bonuses, and rising social inequality [7]. Hence, it is essential to gain insight into and address unmet needs to enhance the effectiveness of KB programs in Indonesia.

The unmet need in East Java in 2023 remains relatively high and constitutes a strategic issue for the local government. Based on BKKBN data, the unmet need was 12.97%, exceeding the target of 11.74%. The figure indicates that the KB service is not yet optimal and that the needs of most fertile couples remain unsatisfied [8]. However, more recent statistics show a changing trend. In 2024, the unmet need rate in East Java decreased to 5.12%, which was already lower than the national target set in the 2020–2024 National Medium-Term Development Plan (RPJMN), namely 7.40%. In 2025, the unmet need rate was recorded at 8.8%, which is still significantly lower than the national average of 13.4%. These variations indicate that although progress has been made, disparities across regions may still exist. Therefore, a data-informed approach and regional mapping are needed to understand spatial differences in unmet need better and to design targeted interventions based on local conditions [9]. Such an approach is essential to ensure that family planning programs are implemented effectively and sustainably across different regions [10].

Fertile Age Couples (PUS) are most likely to experience unmet need as they are in reproductive age and are feeling the desire to postpone or prevent pregnancy, but are not practicing contraception [8]. Socioeconomic status, through this research identified via the poverty index, also influences this significantly since women from low economic status families are more susceptible to unmet need compared to middle-to-upper-level economic classes [9]. The median age at first marriage is an important indicator reflecting the onset of contraceptive needs; the lower the age of marriage, the longer the reproductive period, which ultimately increases the risk of unmet need [11]. Female labor force participation also has implications for fulfilling KB needs, as women's involvement in productive activities can limit opportunities and access to contraceptive services [12]. Unmet need is also an important variable to highlight because it illustrates the extent to which family planning programs can meet community needs, while also identifying existing barriers to the use of contraceptive services. Thus, these five variables are used in this study because they are closely related to the occurrence of unmet need for family planning [2].

Technological advancements offer solutions through the use of clustering methods to group regions based on their specific characteristics and needs, so that the KB services provided are more targeted and effective [13]. Clustering is an important data analysis technique that falls under the category of unsupervised learning, aiming to group data into several clusters based on the similarity of characteristics between objects [14]. One commonly used method, Fuzzy C-Means, is a grouping technique that allows each data point to have a degree of membership in several clusters simultaneously. At the same time, K-Means explicitly groups data points based on their proximity to the cluster centers. Both methods have their respective advantages that can be utilized according to the data characteristics and analysis objectives [15]. The advantage of Fuzzy C-Means is its ability to group overlapping data by assigning degrees of membership to multiple clusters simultaneously, making the results more flexible and better reflecting data uncertainty. K-Means, on the other hand, is faster and simpler, suitable for large datasets, because it produces clear and easily understandable clusters [10].

Research by Nabila Nur Fransiska et al. shows that the K-Means method can produce a Silhouette Coefficient value of 0.576 in grouping poverty data in West Java Province, which falls into the medium structure category and indicates good cluster quality [16]. On the other hand, research by Ifadah et al. on grouping flood-prone areas using the Fuzzy C-Means (FCM) method successfully obtained a Silhouette Coefficient value of 0.84377, indicating a perfect cluster structure [17].

The differences in results indicate that no definitive conclusion can be drawn about which method is superior for mapping unmet need for KB factors. Therefore, this study aims to compare the performance of K-Means and FCM in grouping unmet needs for KB factors in East Java Province. This research aims to provide a spatial overview that is beneficial to local governments in developing KB service strategies and to help identify areas that require special attention. This research contributes to applying a data-based approach to determine more targeted intervention strategies in areas with different levels of unmet need.

II. Research Method

This study uses secondary data from the official websites of Kemendukbangga/BKKBN East Java, covering 38 regencies/cities in 2023. The data used in this study were previously consulted and verified with BKKBN East Java to ensure their accuracy, relevance, and appropriate interpretation of the selected variables and indicators. Data processing in this study was carried out using Python on the Google Colab platform to support practical clustering analysis. This study uses a quantitative approach with cluster analysis to group variables based on their characteristics. The research stages include data collection, data preprocessing (including normalization), application of the Fuzzy C-Means and K-Means clustering algorithms, model evaluation, and data denormalization. Cluster results were evaluated using validation methods such as the Silhouette Coefficient and the Davies-Bouldin Index. The workflow of this research is presented in Figure 1. The variables to be used in this research are shown in Table I.

TABLE I
VARIABLE DESCRIPTION

Variable Code	Independent Variables
X_1	Unmet need for family planning
X_2	Partner Age Fertile
X_3	Index Poverty
X_4	Median Age at First Marriage
X_5	TPAK Women

In this study, five variables related to family planning conditions are used as clustering variables. The Unmet Need for family planning variable reflects the proportion of women whose need for contraception is not met. The Partner Age Fertile (PUS) variable represents the number of couples who potentially need family planning services, where regions with a larger number of fertile couples tend to have higher demand for contraceptive services. The Index of Poverty reflects socioeconomic conditions that may influence access to reproductive health services, as poorer households often face financial and informational barriers in obtaining family planning services. The Median Age at First Marriage is related to the beginning and duration of the reproductive period, where an earlier age at marriage may increase the need for contraception and potentially contribute to a higher level of unmet need. Meanwhile, TPAK Women reflects women's participation in economic activities, which may increase women's autonomy, access to information, and decision-making power related to reproductive health and family planning. Therefore, these variables represent demographic and socioeconomic factors that are associated with the level of Unmet Need for Family Planning across regions.

This study adopts an unsupervised clustering approach, treating all variables as clustering features rather than classifying them as dependent or independent variables. The clustering process aims to group regions based on patterns of similarity across these variables.

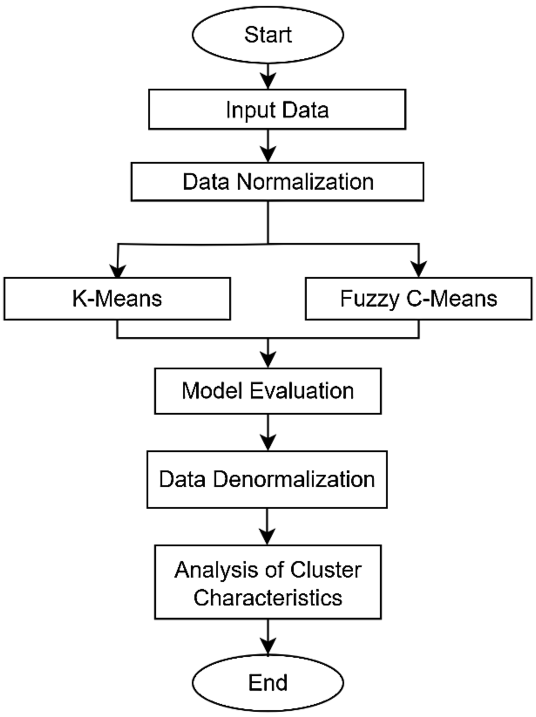

Figure 1. General Research Flow

A. Data Preprocessing

The first step before starting the analysis phase is data preparation. To ensure all variables are on the same scale and to produce more accurate clustering algorithms, preprocessing such as normalization is necessary. However, before normalization, researchers also examine descriptive statistics for each variable, such as minimum, maximum, mean, and standard deviation, to understand the distribution and initial characteristics of the analyzed variables. This is crucial for improving the performance of clustering methods. Normalization is a preprocessing technique that maps or scales data values. This study uses the Min-Max Normalization technique, which is a widely used method for addressing significant differences in feature values. All features are normalized by scaling their values between 0 and 1 to ensure that all data are on the same scale and none dominate the clustering process [18].

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

B. K-Means

The first step in K-Means is to determine the initial centroids and the number of clusters randomly. Next, the distance of each data point to each centroid is calculated using the Euclidean distance formula:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (2)$$

Where x_i is the centroid of object i in the data x, and y_i is the centroid of object i in the data y, and the number of data dimensions is denoted by n. Each data point is then grouped according to its closest centroid. After the initial grouping process is complete, new centroids are calculated by averaging the data in each cluster using the equation.

$$c_k = \frac{1}{n_k} \sum x_i \quad (3)$$

Where n_k is this the number of data points in the cluster k, and x_i is the i-th data point that is a member of the cluster k. Since this process is iterative, it must be ensured that no members move between clusters, and the calculation will stop [10]. The K-Means algorithm was implemented in Python using the scikit-learn library with k = 3 clusters. The initial centroids were randomly selected from the dataset, with seed = 42 to ensure reproducibility. The algorithm was executed with n_init = 1 and a maximum of 100 iterations until convergence was reached.

C. Fuzzy C-Means

The initial step in using the Fuzzy C-Means method for data clustering is to arrange the data into a matrix X sized $n\ x\ j$, where n is the number of data points and j is the number of attributes for each data point. After that, initial calculation values are determined, such as the number of clusters, the exponent value (m), the maximum iteration, and the smallest error. Next, an initial partition matrix U is formed by stating the degree of membership of a data point i in cluster k. Each data point's membership sum must satisfy the following condition.

$$\sum_{k=1}^{c} \mu_{ik} = 1 \quad (4)$$

The second step is to calculate the cluster centers using the following equation.

$$v_k = \frac{\sum_{i=1}^{n} (\mu_{ik})^m \cdot x_i}{\sum_{i=1}^{n} (\mu_{ik})^m} \quad (5)$$

Where μ_{ik} is the degree of membership of the data point i to cluster k, and m is the symbol for the fuzzy exponent value. The third step is to calculate the objective function P_t for the t-th iteration.

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} ||x_i - v_k||^2 \cdot (\mu_{ik})^m \quad (6)$$

The objective function value at the t-th iteration, denoted by P_t measures the total weighted within-cluster variance.

Then, the calculation marks the new membership degrees μ_{ik} is:

$$\mu_{ik} = \left[\sum_{j=1}^{c}\left(\frac{||x_i-v_k||}{||x_i-v_j||}\right)^{\frac{2}{m-1}}\right]^{-1} \quad (7)$$

The next step is to determine whether the smallest error (ε) has been achieved by checking the stopping condition using the maximum iteration formula. The iteration process will be stopped if the maximum number of iterations has been reached.

$$||U^t - U^{t-1}|| < \varepsilon \quad (8)$$

After the iteration process stops, the final step is to cluster the data into the determined clusters [19] . The Fuzzy C-Means algorithm was implemented using the scikit-fuzzy library in Python with a fuzziness exponent m = 2, a stopping tolerance of 0.005, and a maximum of 1000 iterations. The membership matrix was initialized automatically by the algorithm (init = None).

D. Denormalization

Denormalization is the inverse of normalization, restoring normalized data to its original scale. This procedure is essential so that analytical results, such as clustering outcomes or evaluation scores, can be interpreted in their original measurement units and directly related to real-world phenomena. In other words, denormalization enables the processed results to be meaningfully linked to the empirical conditions under investigation.

The primary objective of denormalization is to facilitate accurate interpretation of analytical findings and to support decision-making processes based on actual data values rather than scaled representations. By converting normalized values back to their original range, we can better understand the magnitude and practical implications of the clustering results. The denormalization process based on the Min–Max scaling transformation is presented in Equation (9):

$$d = d'(x_{max} - x_{min}) + x_{min} \quad (9)$$

d' denotes the normalized value, d represents the original value after denormalization, and x_{min} and x_{max} are the minimum and maximum values of the original dataset, respectively.

E. Clustering Model Evaluation

1. *Silhouette Coefficient*

The Silhouette Coefficient indicates how similar each data point is to its own cluster and is calculated independently for each object in the group. When the Silhouette Coefficient approaches 1, the clustering quality within that group increases. However, the closer the cluster is to -1, the worse its clustering quality. In short, the Silhouette Coefficient technique is represented by the following equation:

$$s_i = \frac{b_i - a_i}{(a_i, b_i)} \quad (10)$$

Component a_i is the average distance between the i and all other objects in the same group, while b_i is the minimum value of the average distance of object i to other objects in different groups [20].

2. *Davies Bouldin Index*

The Davies-Bouldin Index is a metric used to evaluate clustering methods internally, based on cohesion and separation. In the context of clustering, cohesion refers to the closeness of data to their respective cluster centroids, while separation is the distance between cluster centroids. When the distance within a cluster is relatively small but significant, all objects have a high degree of feature similarity. A desirable clustering structure is characterized by low intra-cluster dispersion and high inter-cluster separation. [21]. The DBI formula is described as follows.

$$R_{i,j} = \frac{S_i + S_j}{d(M_i, M_j)} \quad (11)$$

The calculation of the quality of clustering results $R_{i,j}$ is presented as follows, where the average distance between each data point in cluster i and j is denoted by $S_i \; dan \; S_j$, while $d(M_i, M_j)$ represents the Euclidean distance between cluster centers.

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i} R_{i,j} \quad (12)$$

Where k represents the number of clusters used, and $R_{i,j}$ is the ratio value from the result of Equation 11.

III. Results and Discussion

This section presents the results of the clustering analysis of factors influencing unmet family planning need in East Java Province in 2023.

TABLE II
DATA ON UNMET NEED FACTORS

Regency/City	X_1	X_2	X_3	X_4	X_5
Pacitan	8.88	14.87	0.25	19.9	73.91
Ponorogo	61.23	1.41	0.17	22.7	64.23
Trenggalek	9.64	16.02	0.28	19.9	72.45
⋮	⋮	⋮	⋮	⋮	⋮
Batu City	12.52	15.68	0.8	21.2	69.48

Based on Table II, we analyzed 38 regencies/cities in East Java Province with ratio data consisting of five indicators: unmet need for family planning (X_1), fertile age couples (X_2), poverty index (X_3), median age at first marriage (X_4), and female labor force participation rate (X_5). In analyzing the data, it is essential to understand the data's initial condition. One way to recognize data characteristics is through descriptive statistical analysis. Table III shows the descriptive statistics for each variable.

TABLE III
DESCRIPTIVE STATISTICS

Variables	Min	Mean	Max	Standard Deviation
X_1 (%)	0.68	15.27	84.68	17.14
X_2 (%)	1.37	13.19	17.43	4.34
X_3 (%)	0.06	0.34	1.53	0.29
X_4	18.20	21.02	24.00	1.50
X_5 (%)	48.86	60.75	73.91	6.02

Based on the data in Table III, the average unmet need for Family Planning (KB) was recorded at 15.27%, with a standard deviation of 17.14%. The average proportion of fertile-age couples (PUS) was 13.19%. This variation is also reflected in a standard deviation of 4.34% of couples. Furthermore, the poverty index averages 0.34, with a standard deviation of 0.29, indicating significant disparities in poverty levels in the studied regions. For the median age at first marriage of women, the average was 21.02 years, with a range of 18.2 to 24 years. A standard deviation of 1.50 years indicates relatively moderate variation in the age at first marriage among women across regions. Meanwhile, the female labor force participation rate (TPAK) averages 60.75%, ranging from 48.86% to 73.91%. The standard deviation of 6.02% indicates differences in the level of women's workforce participation in the research area. The researchers presented the following table to identify regions with minimum and maximum values.

Analysis of extreme values per variable in Table III shows that Situbondo has the lowest unmet family planning need at 0.68%, indicating low demand but possibly limited access to services. Madiun has the lowest PUS at 1.37%, which can affect KB programs. Although inequality may still exist, Banyuwangi recorded the lowest poverty index at 0.06. Nganjuk has the lowest female labor force participation rate at 48.86%, and Bondowoso has the lowest median age at first marriage at 18.20 years, potentially increasing the incidence of early marriage. On the other hand, Madiun has the highest unmet need for family planning at 84.68%, indicating high unmet services and a higher risk of family health problems. Although this value appears substantially higher than the overall mean, it was retained in the analysis because it represents official regional statistics and reflects actual variation among districts, thereby preserving the dataset's representativeness. The highest PUS is in Probolinggo at 17.43%, and Sumenep has the highest poverty index at 1.53, indicating extreme social inequality. While Pacitan has the highest female labor force participation rate at 73.91%, indicating high female participation, Kota Madiun has the highest median age at first marriage at 24 years, indicating more mature marriages. However, both must be supported by balanced policies.

This indicates a significant disparity between regions regarding unmet family planning needs. Before applying the clustering method, data normalization was performed to ensure equal scaling across the analyzed variables. Table IV presents the normalization results for each indicator used in this study.

TABLE IV
NORMALIZATION TABLE

Regency /City	X_1	X_2	X_3	X_4	X_5
Pacitan	0.0976	0.8405	0.1292	0.2931	1.0000
Ponorogo	0.7208	0.0025	0.0748	0.7759	0.6135
Trenggalek	0.1067	0.9122	0.1497	0.2931	0.9417
⋮	⋮	⋮	⋮	⋮	⋮
Batu City	0.1410	0.8910	0.5034	0.5172	0.8232

After normalization, the next step was to cluster the data. Then, the optimal number of clusters was determined using the Silhouette Coefficient and the Davies-Bouldin Index.

A. *Optimal Cluster Determination*

Model evaluation in this study was conducted to determine the optimal number of clusters and to compare the performance of two clustering methods, namely K-Means and Fuzzy C-Means. The optimal cluster determination used two evaluation methods: Silhouette Coefficient and Davies Bouldin Index (DBI). The determination of the number of clusters within the range of 2 to 5 was conducted to maintain a balance between analytical depth and interpretability. A number of clusters allows the grouping results to remain representative of data variability while still being clear and relevant for further analysis.

TABLE V
COMPARISON RESULTS OF SILHOUETTE COEFFICIENT AND DBI

Cluster	Silhouette Coefficient		Davies Bouldin Index	
	K-Means	Fuzzy C-Means	K-Means	Fuzzy C-Means
2	0.6903	0.5326	0.4732	0.5298
3	**0.6997**	0.2447	**0.3539**	1.2678
4	0.5860	0.2259	0.3458	12.195
5	0.4791	0.2044	0.6033	15.029

Based on the results presented in Table V, K-Means with three clusters produced the highest Silhouette Coefficient value (0.6997) and the lowest DBI value (0.3539). These results indicate that the three-cluster solution yielded compact, well-separated clusters. Therefore, K-Means with three clusters was selected as the optimal clustering model in this study.

Fuzzy C-Means yielded lower Silhouette values and higher DBI values across most cluster scenarios. The best performance of Fuzzy C-Means was observed at $k = 2$, with a Silhouette value of 0.5326 and a DBI of 0.5298, which were still inferior to the corresponding K-Means results. This suggests that the clusters formed by Fuzzy C-Means were less compact and less distinctly separated.

The superior performance of K-Means may be explained by the dataset's structure, which consists of aggregated regional socioeconomic indicators across 38 regencies/cities. Such data tend to exhibit relatively distinct regional profiles rather than overlapping membership patterns. In this context,

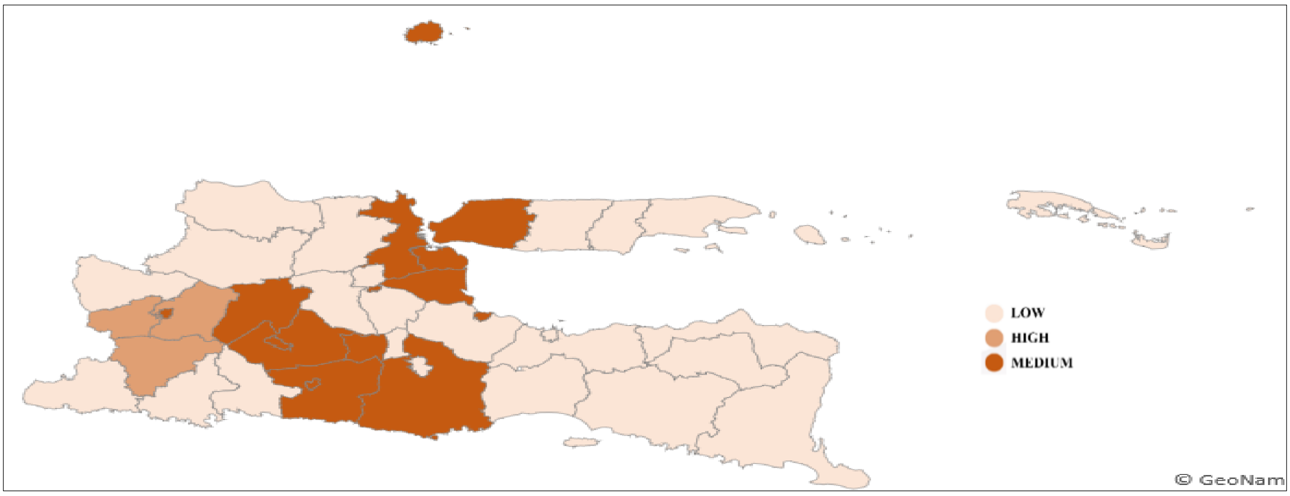

Figure 2. Map of K-Means Clustering Results Distribution

a hard clustering approach is more appropriate, as it assigns each region exclusively to a single cluster, thereby enhancing interpretability and stability [10]. Conversely, Fuzzy C-Means is generally more suitable for datasets characterized by gradual transitions and high ambiguity between groups [15]. Since the current dataset does not exhibit strong boundary overlap, the fuzzy membership mechanism did not provide additional benefits.

After selecting the optimal clustering model, the normalized data were denormalized to their original scale to facilitate meaningful interpretation of cluster characteristics. Denormalization was performed using Equation (9), and the results are presented in Table VI.

TABLE VI
DATA AFTER DENORMALIZATION

Regency/City	X_1	X_2	X_3	X_4	X_5
Pacitan	8.88	14.87	0.25	19.9	73.91
Ponorogo	61.23	1.41	0.17	22.7	64.23
Trenggalek	9.64	16.02	0.28	19.9	72.45
⋮	⋮	⋮	⋮	⋮	⋮
Batu City	12.52	15.68	0.8	21.2	69.48

Overall, the evaluation results indicate that the K-Means algorithm provides better clustering performance than Fuzzy C-Means for grouping regions based on unmet need for family planning indicators in East Java Province, especially with 3 clusters, resulting in higher clustering quality. The cluster index for each regency/city based on the K-Means algorithm results is presented below.

TABLE VII
CLUSTER LABELS

Regency/City	Cluster
Pacitan	0
Ponorogo	1
Trenggalek	0
Tulungagung	0
Blitar	2
⋮	⋮
Batu City	0

The detailed grouping of unmet family planning needs in East Java using the K-Means algorithm with the optimal number of clusters k = 3 is shown in Table VII.

TABLE VIII
K-MEANS CLUSTERING RESULTS

Cluster	Number of Regions	Member Group
1	22	Pacitan, Trenggalek, Tulungagung, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Mojokerto, Jombang, Ngawi, Bojonegoro, Tuban, Lamongan, Sampang, Pamekasan, Sumenep, Probolinggo City, Batu City
2	3	Ponorogo, Madiun, Magetan
3	13	Blitar, Kediri, Sidoarjo, Nganjuk, Gresik, Bangkalan, Kediri City, Blitar City, Malang City, Pasuruan City, Mojokerto City, Madiun City, Surabaya City

Based on the K-Means clustering results presented in Table VIII, three regions in East Java were identified as clusters. To understand the characteristics of each cluster after denormalization, refer to Table IX, which displays the average variable values.

TABLE IX
AVERAGE VARIABLES CLUSTER

Variables	Cluster			Average
	1	2	3	
Unmet need for family planning (%)	8.79	70.52	13.47	15.27
Partner Age Fertile (%)	15.19	1.39	12.58	13.19
Index Poverty (%)	0.40	0.21	0.26	0.34
Median Age at First Marriage	19.98	22.53	22.43	21.02
TPAK Women (%)	61.94	64.24	57.91	60.75

Based on Table IX, the clusters are interpreted according to the level of unmet family planning need, while the other variables serve as supporting indicators of each cluster's socioeconomic conditions. Cluster 1 has an unmet need for family planning of 8.79, which is lower than the overall average of 15.27, indicating that the demand for family planning services in this cluster is relatively low. This condition suggests that the need for family planning among fertile-age couples (PUS) in this cluster is more met than in other clusters. However, the poverty index is relatively high (0.40), and the median age at first marriage is the lowest (19.98) among all clusters, indicating the presence of socioeconomic challenges and early marriage patterns. These factors may influence reproductive behavior and the future demand for family planning services in this cluster.

Cluster 2 shows a very high unmet need for family planning of 70.52, well above the overall average of 15.27. A high unmet need indicates that a large proportion of fertile-age couples who wish to delay or limit pregnancy are not using contraceptive methods, meaning that the demand for family planning services remains unmet. Although this cluster has relatively favorable socioeconomic indicators, such as the lowest poverty index (0.21), the highest female labor force participation rate (64.24), and the highest median age at first marriage (22.53), the extremely high unmet need may indicate the presence of non-economic barriers. These barriers may include limited access to family planning services, a lack of information about contraceptive methods, or social and cultural factors influencing contraceptive use.

Cluster 3 has an unmet need for family planning of 13.47, which is close to the overall average of 15.27, indicating a moderate level of unmet need. The number of fertile-age couples (12.58) and the poverty index (0.26) are also at moderate levels compared to other clusters. In addition, the median age at first marriage (22.43) indicates relatively mature marriage patterns that may support better reproductive planning. However, the female labor force participation rate (57.91) is the lowest among the clusters. Overall, this cluster has relatively moderate family planning indicators compared to the other clusters.

The mapping results show that East Java is divided into three clusters based on the level of unmet need for family planning: Cluster 1 with the lowest level of unmet need, indicating relatively lower demand for family planning services; Cluster 2 with the highest level of unmet need, suggesting that a large proportion of fertile-age couples who wish to delay or limit pregnancy are not using contraceptive methods; and Cluster 3 with a moderate level of unmet need close to the overall average. This clustering pattern highlights regional variation in family planning needs. It suggests that Cluster 2 should be prioritized for improving access to contraceptive services and reproductive health education. In contrast, Cluster 1 and Cluster 3 require preventive and supportive strategies tailored to their respective socioeconomic characteristics.

The spatial distribution of clusters shows a clear geographical pattern across East Java. Cluster 2, which represents regions with the highest level of unmet need for family planning, is predominantly located in the western part of East Java. This spatial concentration indicates that several neighboring districts share similar characteristics related to unmet family planning needs. Meanwhile, Cluster 1, characterized by the lowest level of unmet need, is more widely distributed in the eastern part of the province. Cluster 3, which represents a moderate level of unmet need, is mostly found in the central and southern areas. This spatial pattern suggests that geographic proximity may influence similarities in socioeconomic conditions and access to family planning services among districts.

Despite these findings, this study has several limitations. The analysis relies on a limited number of variables derived from secondary data. It does not fully capture broader social, cultural, and behavioral dynamics that may influence unmet need for family planning across regions.

IV. CONCLUSION

This study concludes that the level of unmet need for family planning in East Java in 2023 was successfully grouped into three regional clusters using the K-Means method: the first cluster comprising 20 regencies and two cities; the second, three regencies; and the third, 6 regencies and 7 cities. The K-Means model's performance evaluation, with a Silhouette Coefficient of 0.6997 and a Davies–Bouldin Index of 0.3539, indicates a compact, well-separated clustering structure. Therefore, the K-Means method is recommended for grouping unmet needs for family planning data in East Java. Future research is suggested to integrate spatial analysis with clustering methods to identify geographical patterns of unmet need for KB and add variables related to unmet need and the distribution of health resources to support more targeted policy recommendations.

REFERENCES

[1] S. K. Sejati, “Analisis Cluster Unmet Need Keluarga Berencana Di Indonesia,” *JLS*, vol. 4, no. 2, p. 10, 2020, doi: 10.32630/sukowati.v4i2.158.

[2] B. D. Saputri and D. Indriani, “Pemetaan Cakupan Pengguna Kb Aktif Dan Unmet Need Di Kabupaten Ponorogo Tahun 2021,” *Jurnal Statistika*, vol. 15, no. 1, 2022, doi: 10.36456/jstat.vol15.no1.a5446.

[3] A. K. Azizah, F. M. Azkiya, and A. P. Puteri, “Implementasi Kebijakan Program Keluarga Berencana Di Puskesmas Kedungdoro Kecamatan Tegalsari Kota Surabaya,” *INTELLECTUAL*, vol. 10, no. 1, pp. 110–118, 2023, doi: 10.55499/intelektual.v10i1.1045.

[4] D. A. Mukti and A. Prabawa, “Autokorelasi Spasial Unmet Need Keluarga Berencana di Pulau Sumatera Tahun 2022,” *Media Publikasi Kesehatan Promosi Indonesia*, vol. 7, no. 5, pp. 1220–1225, 2024, doi: 10.56338/mppki.v7i5.5060.

[5] S. Nisrina, W. P. Nurmayanti, Basirun, Kertanah, and M. Gazali, “Penerapan Metode Clustering SOM dan DBSCAN dalam Mengelompokkan Unmet Need Keluarga Berencana di Nusa Tenggara Barat,” *Jurnal Statistika*, vol. 15, no. 2, pp. 237–244, 2022, doi: 10.36456/jstat.vol15.no2.a5549.

[6] A. P. Pramudiyanti, M. F. Shafiro, L. A. Salim, and Wasyik, “K-Means Cluster Analysis Related To Unmet Need For Family Planning In Banyuwangi, Indonesia: A Case Study,” *JPHRECODE*, vol. 7, no. 2, pp. 136–142, 2024, doi: 10.20473/jphrecode.v7i2.39691.

[7] G. R. Suraya and A. W. Wijayanto, “Comparison of Hierarchical Clustering, K-Means, K-Medoids, and Fuzzy C-Means Methods in Grouping Provinces in Indonesia according to the Special Index for Handling Stunting,” *Indonesian Journal of Statistics and Applications*, vol. 6, no. 2, pp. 180–201, 2022, doi: 10.29244/ijsa.v6i2p180-201.

[8] Sumiyati, I. Kurniawan, A. Hakim, C. Ermanto, and A. Ali, “Faktor-Faktor yang Mempengaruhi Unmet Need (Analisis Berdasar Hasil Capaian Data Pemutakhiran Keluarga Tahun 2023),” *Action Research Literate*, vol. 8, no. 7, Jul. 2024.

[9] Kusmanto, S. Samsir, R. Watrianthos, and S. Suryadi, “Distribusi Spasial Unmet Need Pelayanan Kesehatan dengan Algoritma K-Means untuk Pemetaan Provinsi di Indonesia,” *Bulletin of Information Technology*, vol. 4, no. 3, pp. 362–368, 2023, doi: 10.47065/bit.v4i3.862.

[10] T. G. Manik, W. I. Rahayu, and Rd. N. S. Fathonah, “Perbandingan Metode Fuzzy C-Cmeans dan K-Means Clustering pada Data Penggunaan Obat di R.S National Hospital Surabaya,” *JATI*, vol. 7, no. 1, pp. 591–597, 2023, doi: 10.36040/jati.v7i1.6430.

[11] A. Zulfikar and F. A. Rahmani, “Deteksi Anomali Menggunakan Isolation Forest Belanja Barang Persediaan Konsumsi pada Satuan Kerja Kepolisian Republik Indonesia,” *Jurnal Manajemen dan Perilaku*, vol. 4, no. 1, pp. 1–15, 2023, doi: 10.33105/jmp.v4i1.435.

[12] S. Sariyati, S. Mulyaningsih, and S. Sugiharti, “Faktor yang Berhubungan dengan Terjadinya Unmet Need KB pada Pasangan Usia Subur (PUS) di Kota Yogyakarta,” *JNKI*, vol. 3, no. 3, pp. 123–128, 2016, doi: 10.21927/jnki.2015.3(3).123-128.

[13] D. Sulistiawan, E. Gustina, R. Matahari, and V. Marthasari, “Profil Sosiodemografis Unmet Need Keluarga Berencana pada Wanita Kawin di Daerah Istimewa Yogyakarta,” *Jurnal Kesehatan Berbasis Kebijakan*, vol. 5, no. 2, pp. 1–9, 2021, doi: 10.37306/kkb.v5i2.49.

[14] A. Kurnia, “Perbandingan Algoritma K-Means dan Fuzzy C-Means untuk Clustering Puskesmas Berdasarkan Gizi Balita di Surabaya,” *Processor*, vol. 18, no. 1, 2023, doi: 10.33998/processor.2023.18.1.696.

[15] R. Nur’aini, T. Widiharih, and B. A. Saputra, “Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Kesehatan Bayi dan Balita Menggunakan Algoritma Fuzzy C-Means dan K-Medoids,” *Jurnal Gaussian*, vol. 13, no. 1, pp. 189–198, 2024, doi: 10.14710/j.gauss.13.1.189-198.

[16] Badan Kependudukan dan Keluarga Berencana Nasional Provinsi Jawa Timur, *Profil Kependudukan Provinsi Jawa Timur*. BKKBN Jawa Timur, 2023.

[17] N. N. F. R, D. S. Anggraeni, and U. Enri, “Pengelompokkan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient,” *THEMATIC*, vol. 9, no. 1, pp. 29–35, 2022, doi: 10.38204/tematik.v9i1.901.

[18] S. Haviyola, S. Susilawati, and M. Jajuli, “Pengelompokan Prestasi Siswa Guna Kualifikasi Beasiswa Berdasarkan Data Nilai Menggunakan Algoritma K-Means,” *JATI*, vol. 7, no. 4, pp. 2786–2791, 2024, doi: 10.36040/jati.v7i4.7200.

[19] S. Dwididanti and D. A. Anggoro, “Analisis Perbandingan Algoritma Bisecting K-Means dan Fuzzy C-Means pada Data Pengguna Kartu Kredit,” *EMITOR*, vol. 22, no. 2, pp. 110–117, 2022, doi: 10.23917/emitor.v22i2.15677.

[20] R. N. Fadhila, N. Ulinnuha, and M. Hafiyusholeh, “Analysis of Inflation Rates During and After the COVID-19 Pandemic Using the K-Means Clustering Method and Kruskal-Wallis Test,” vol. 14, no. 2, pp. 56–67, 2025, doi: 10.14421/fourier.2025.142.56-67.

[21] Muhammad Raqib Syahkur, D. Hartama, and S. Solikhun, “Evaluasi Jumlah Cluster pada Algoritma K-Means++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita,” *JST (Jurnal Sains dan Teknologi)*, vol. 13, no. 3, pp. 487–496, Oct. 2024, doi: 10.23887/jstundiksha.v13i3.86419.