Jurnal CoreIT, Vol.11, No.2, December 2025
ISSN 2460-738X (Print)
ISSN 2599-3321 (Online)

Data Mining for Analyzing Consumer Segmentation: Identifying Consumer Preference Patterns Using the Fuzzy C-Means Clustering on Halal Products

Iwan Iskandar[1]*, Alwis Nazir[1]
[1]Dept. of Informatics Engineering, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
iwan.iskandar@uin-suska.ac.id*, alwis.nazir@uin-suska.ac.id

Abstract. Halal products are increasingly popular worldwide, not only in Muslim-majority countries but also in non-Muslim nations. The global halal market exceeds USD 650 million annually, emphasizing the importance of halal certification, particularly in Indonesia as the world's largest Muslim-majority country. This research aims to cluster consumers of halal meat products by analyzing factors influencing consumer behavior in purchasing certified halal beef and chicken. The study employs the Fuzzy C-Means (FCM) clustering algorithm on 176 respondents' questionnaire data consisting of 36 parameters. The experiment was performed using Google Colab with a maximum of 1000 iterations, error tolerance of 0.0001, and fuzziness coefficient m = 2.4. Results show that two optimal clusters were formed, with a Partition Coefficient Index (PCI) value of 0.9993, indicating excellent clustering quality. The first cluster consists primarily of young consumers aged 15–24 with lower spending, while the second cluster includes adults aged 35–54 with higher income. Both groups prioritize halal certification and logo visibility when choosing meat products. The findings provide insights for halal product retailers and policymakers to enhance halal product distribution, certification support, and marketing strategies in Indonesia.

Keywords : Clustering, Fuzzy C-Means, Halal Product, Consumer Behavior

Received November 2025 / **Revised** December 2025 / **Accepted** December 2025

INTRODUCTION

Halal products have become increasingly popular worldwide, not only in Muslim-majority countries but also in many other parts of the world. Transactions involving halal products, valued at more than USD 650 million, occur annually, reflecting a growing global demand. According to *The Royal Islamic Strategic Studies Centre (RISSC)* in its report *"The Muslim 500: The World's 500 Most Influential Muslims 2024"*, Indonesia ranks as the country with the largest Muslim population in the world. The report recorded that in 2023, Indonesia's Muslim population reached 240.62 million people, equivalent to 86.7% of the total national population of 277.53 million [1]. As a nation with a Muslim-majority population, Indonesia should prioritize the needs of its citizens regarding the consumption of halal products, including ensuring the halal assurance of the products they consume.

For Muslims, consuming halal food is a religious obligation, as commanded by Allah in *Surah Al-Baqarah (2:88)*, *Al-Maidah (5:88)*, *Al-A'raf (7:157)*, and *Al-An'am (6:188)*. The management and supervision of halal products in Indonesia are currently regulated by the *Halal Product Assurance Organizing Agency* (*Badan Penyelenggara Jaminan Produk Halal*, BPJPH), established under *Law No. 33 of 2014 on Halal Product Assurance* [2]. BPJPH is responsible for halal certification and supervision throughout Indonesia, in collaboration with the *Halal Inspection Agency* (LPH) and the *Indonesian Ulema Council* (MUI) as the issuer of halal fatwas. In addition, the law provides support for micro-entrepreneurs through access to funding from the *State Budget* (APBN) and *Regional Budgets* (APBD) to facilitate the halal certification process.

Consumer decisions regarding product distribution channels depend on their preferences, financial capacity, and the type of product desired. The consumption of beef and chicken continues to rise in Indonesia. Based on data from the *Central Bureau of Statistics* (BPS), the average consumption of beef/buffalo meat in 2022 was 0.010 kg per capita per week, while the production of broiler chicken in Riau reached 104,331.48 tons during the same year [3]. Meanwhile, the demand for frozen meat products also increased, totaling 153,637,500 tons for red meat and 376,275,000 units for poultry chicken [4].

DOI: 10.24014/coreit.v11i2.38608

This study focuses on clustering consumers by analyzing factors influencing their behavior in selecting certified-halal beef/buffalo and chicken products. These two commodities were chosen because they represent the main protein sources among animal-based foods. The *Fuzzy C-Means (FCM)* clustering method is employed due to its high accuracy in identifying inter-cluster relationships [5] – [7]. A previous study conducted by Deviana et al., titled *"Application of Fuzzy C-Means Method for Grouping Regencies/Cities on the Island of Kalimantan Based on the 2020 Human Welfare Indicators,"* demonstrated that the Fuzzy C-Means method can produce optimal clusters in regional groupings based on welfare levels [8].

The expected outcome of this study is to identify consumer preference patterns and classify consumers who purchase halal products based on those patterns. The results are expected to provide valuable insights for halal product producers to analyze sales growth and for policymakers to design more effective future strategies and regulations regarding halal certification in Indonesia.

FUZZY C-MEANS CLUSTERING

Fuzzy C-Means (FCM) is an unsupervised learning method used to group data into clusters based on feature similarity. The goal is to minimize the variance within clusters while maximizing data membership accuracy toward the appropriate cluster centers [5],[9],[10].

FCM applies the concept of *fuzzy logic*, which allows each data point to have a partial membership in multiple clusters. Consequently, each data point is assigned a *membership weight* indicating how closely it belongs to each cluster center [11]-[14].

Fuzzy C-Means Algorithm [15],[16] :

1) Initialization: Select the desired number of clusters (*C*) and randomly initialize cluster centers.
2) Membership Calculation: Compute the membership degree (μik) of each data point for each cluster based on the distance between the data point and the cluster center.
3) Cluster Center Update: Update cluster centers by calculating the weighted mean of all data points, using membership values as weights.
4) Iteration: Repeat steps (2) and (3) until convergence is achieved or the maximum number of iterations is reached.

Given a dataset X, represented as a matrix of size $n \times m$ (where n = number of data samples and m = number of attributes), each data point Xij represents the *i-th* sample ($i = 1, 2, \ldots, n$) and *j-th* attribute ($j = 1, 2, \ldots, m$).

Parameters:

C = number of clusters
W = fuzziness exponent
MaxIter = maximum number of iterations
ε = minimum acceptable error
$P_0 = 0$ = initial objective function value
$t = 0$ = initial iteration

Step 1: Generate random values for μik as the initial partition matrix U, satisfying:

$$\sum_k \mu ij = 1 \quad (1)$$

Step 2: Calculate the cluster center *Vkj* for each cluster *k* and attribute *j*:

$$V_{kj} = \frac{\sum_{i=1}^{n} (\mu_{ik})^w \, x_{ij}}{\sum_{i=1}^{n} (\mu_{ik})^w} \quad (2)$$

Step 3**:** Compute the objective function at iteration *t*:

$$P_{t=} \sum_{i=1}^{n} \sum_c [\sum_m (X_{ij-} V_{kj})^2 \ (\mu_{ik})^w \quad (3)$$

After the clustering process, the results are validated using the Partition Coefficient Index (PCI)**,** which evaluates the quality of the clustering results.The PCI assesses the degree of membership in each cluster without considering the actual data values that contain the information. The PCI formula is as follows:

$$PCI = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{C} (\mu_{ij})^2 \quad (4)$$

METHOD

This study was conducted through several sequential stages, including data collection, feature selection, data preprocessing, data transformation, clustering analysis, and result evaluation.

A. Data Collection

Data were collected through a questionnaire distributed to 176 respondents who have either purchased or consumed halal processed chicken and beef products in the form of frozen food. Respondents were consumers of halal-labeled meat or chicken products from supermarkets or frozen food outlets. Each questionnaire used a 5-point Likert scale across ten behavioral constructs (e.g., purchase intensity, subjective norms, price, sensory appeal). Before distribution, the questionnaire was validated to ensure its accuracy, relevance, proper language use, and alignment with the research objectives. The questionnaire consists of various questions divided into 10 categories, covering respondent biodata and consumer performance measurement, including: Purchase Intensity, Consumer Attitude When Purchasing, Subjective Norms, Behavioral Control, Supermarket Location, Product Convenience, Brand Loyalty, Halal Logo, Sensory Appeal and Price. The collected data was then processed to gain insights from each cluster that is formed.

B. Feature Selection

At this stage, a parameter selection process was carried out to identify the most relevant variables for the study. A total of 39 parameters were initially obtained from the questionnaire; however, after a systematic screening process, 36 parameters were selected for further analysis. The retained parameters were those deemed relevant and directly aligned with the objectives and scope of the research. The selected parameters are those that are relevant to the research, including: Gender, Age, Religion, Last Education, Occupation, Monthly Food Expenses, Monthly Income, Frequent Retailers Visited, Favorite Type of Processed Products Consumed, Frequency of Purchasing Processed Chicken or Beef per Week, Frequency of Purchasing Processed Chicken or Beef per Month, Attention to Halal Labels When Purchasing Processed Chicken or Beef, The Importance of Halal Certification in Purchasing Processed Chicken or Beef Over Other Factors, The Requirement of a Halal Label on Products I Purchase, Comfort When Buying Processed Chicken or Beef in Supermarkets, Family Influence on My Consumption of Processed Chicken and Beef Products, Religious Leaders' Influence on My Decision to Buy Halal-Certified Food, Ease of Purchasing Processed Chicken in Supermarkets Compared to Traditional Markets, Confidence in Choosing Halal-Certified Food Over Non-Certified Products, The Influence of Price on My Decision to Buy Halal-Certified Processed Meat, The Influence of Promotions on My Decision to Buy Halal-Certified Processed Meat, Buying Processed Chicken and Beef in Supermarkets Near My Residence, Buying Processed Chicken and Beef in Supermarkets with Comfortable Spaces Even If They Are Far from My Residence, Purchasing Processed Chicken and Beef in Supermarkets with Ongoing Promotions Even If They Are Distant, Inconvenience in Serving Processed Chicken or Beef, The Importance of Convenience in Using Halal-Certified Processed Chicken or Beef, Buying Processed Chicken or Beef from Specific Brands, Only Purchasing Halal-Certified Processed Chicken or Beef from Certain Brands, The Importance of the Halal Logo on Processed Chicken or Beef Packaging, Checking for the Halal Logo Before Purchasing Processed Meat, The Superior Taste of Processed Chicken or Beef Purchased in Supermarkets Compared to Other Brands, The Tender and Chewy Texture of Processed Chicken or Beef Purchased in Supermarkets, The Attractive Packaging of Processed Chicken or Beef Purchased in Supermarkets, The Influence of Visual Appeal (Color, Appearance) of Halal-Certified Processed Chicken or Beef on My Decision to Buy, The Impact of Price on My Decision to Buy Halal-Certified Processed Chicken and Beef, and Purchasing Processed Chicken or Beef Only When On Sale or Discounted.

C. Data Preprocessing

At this stage, a data cleaning process was performed to eliminate inconsistencies such as missing values, duplicate entries, and noisy data. Based on the questionnaire results, all collected data were found to be complete and consistent; therefore, no records were removed from the dataset. This indicates that the data quality was adequate for subsequent analytical processes.

D. Transformation

At this stage, data transformation was conducted for several parameters to be used in subsequent calculations. Categorical variables were converted into numerical representations to facilitate computational analysis. For instance, the gender variable was transformed by assigning a value of 1 to males and 2 to females.

Furthermore, features represented as ordinal ranges, such as age, were discretized into numerical categories. Specifically, children aged 0–14 years were assigned a value of 1, adolescents aged 15–24 years were assigned a value of 2**,** young adults aged 25–34 years were assigned a value of 3**,** adults aged 35–54 years were assigned a value of 4**,** middle-aged individuals aged 55–64 years were assigned a value of 5, and elderly individuals aged 65 years and above were assigned a value of 6.

```
# preview the dataset
data.head()
```

	A	B	C	D	E	F	G	H	I	X1	...	X18	X19	X20	X21	X22	X23	X24	X25	X26	X27
0	1	4	1	4	2	7	6	1	2	3	...	4	3	5	4	4	3	4	4	4	2
1	2	2	1	2	1	2	1	3	3	3	...	3	3	4	4	3	4	4	4	4	3
2	2	2	1	2	1	2	1	2	2	4	...	4	4	4	4	4	3	2	3	3	2
3	1	2	1	2	1	2	1	1	2	5	...	4	5	5	5	5	5	5	5	5	2
4	1	2	1	2	1	2	3	3	2	4	...	4	4	4	4	4	4	4	4	4	4

5 rows × 36 columns

```
[ ] # view dimensions of dataset
    data.shape
```

```
(176, 36)
```

Figure 1. Transformation data and variabel

The same transformation approach was applied consistently to all other features used in the analysis. This process ensured that the dataset was fully numerical and suitable for clustering and further analytical procedures.

Implementation of Fuzzy C-Means clustering

The data is then processed using the Fuzzy C-Means method to generate clusters. Before this, an optimization step is performed on the number of parameters to determine the most suitable dimensions. This optimization process employs the Principal Component Analysis (PCA) method.

```
# 1. Scale the data:
# Create a StandardScaler object

scaler = StandardScaler()

# Fit the scaler to your data and transform it
X_scaled = scaler.fit_transform(data)

# 2. Apply PCA:
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_scaled)

# 3. Print explained variance:
print("Explained Variance Ratio:", pca.explained_variance_ratio_)
```

```
Explained Variance Ratio: [0.17457817 0.12810533]
```

Figure 2. Optimization with PCA method

The calculations yield Explained Variance Ratios (EVR) of 0.1745 and 0.1281. The first component (EVR = 0.1745) explains 17.45% of the total variance in the data, while the second component (EVR = 0.1281) accounts for 12.81% of the total variance.

Subsequently, the results from PCA are used as input data for the clustering process with the Fuzzy C-Means method. The initial phase of initialization uses 2 clusters, with a fuzziness parameter m = 2.4, a maximum of 1000 iterations, and an error tolerance of 0.0001.

```
c = 2  # number of clusters
w = 2.4  # power of numbers
maxIter = 1000  # maximum iterations
error = 0.0001  # smallest error
```

```
cntr, u, u0, d, jm, p, fpc = fuzz.cluster.cmeans(X.T, c=c, m=w, error=error, maxiter=maxIter, init=None)

# Get the cluster membership probabilities
labels = np.argmax(u, axis=0)
```

```
print("Number of iterations:", p)
```

```
Number of iterations: 152
```

Figure 3. Fuzzy C-Means

The results of clustering using Fuzzy C-Means yielded two clusters: Cluster 0 and Cluster 1.

	0	1	cluster_FCM
0	4.170386	0.473243	0
1	-3.625075	-1.706832	1
2	-3.552525	-1.895345	1
3	-3.043699	4.451448	1
4	-2.479498	1.731913	1
...	...	...	...
171	3.826858	0.586136	0
172	0.885471	3.933226	0
173	0.329420	-0.803092	0
174	2.204470	-0.103194	0
175	5.447371	0.354403	0

176 rows × 3 columns

Figure 4. Result of clustering

Figure 5 below illustrates the distribution of each data point according to its respective cluster

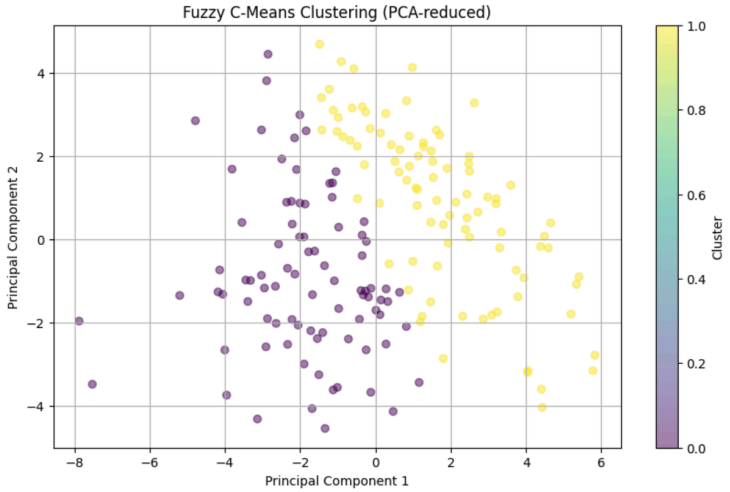

Figure 5. Distribution of each data

RESULTS AND DISCUSSION

The evaluation of this research was conducted using the Partition Coefficient Index (PCI) method. The testing procedure was carried out under five different scenarios. Each scenario used 1000 iterations with an error threshold of 0.0001, while varying the value of *m*, specifically *m* = 2.0, 2.1, 2.2, 2.3, and 2.4, for a total of two clusters. Table 1 presents the PCI results obtained for each scenario based on the different *m* values

Table 1. Evaluation results for each cluster

M VALUE	PCI
2,0	0.6866064050033187
2,1	0.7331647716152218
2,2	0.7905572178671754
2,3	0.8755748437603061
2,4	0.9993799515177483

From table 1, it can be concluded that the highest PCI value is achieved at *m* = 2.4, which is 0.9993799515177483. Subsequently, testing was conducted to determine the most optimal number of clusters. In this study, the testing was performed for the number of clusters ranging from 2 to 10. The results indicate that the 2-cluster configuration yields the highest PCI value. Figure 6 below illustrates the optimal values for *K* = 2 through *K* = 10.

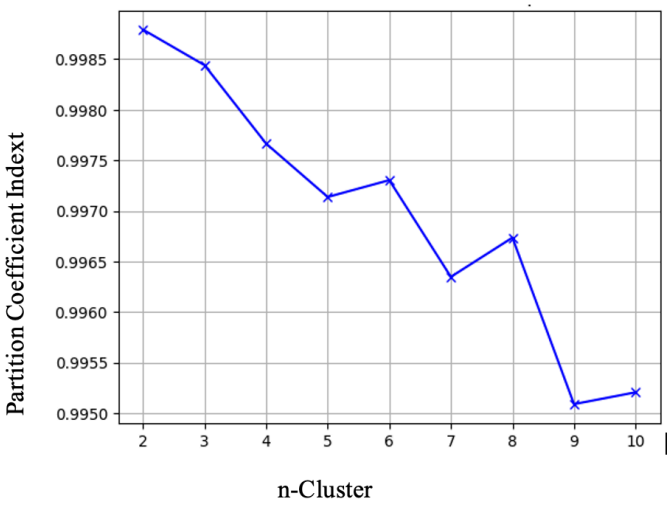

Figure 6. K-Optimum

Based on the test results, it was found that the optimal clustering solution consisted of two clusters, namely Cluster One (Cluster 0) and Cluster Two (Cluster 1). From these two clusters, several insights can be derived, as described below:

1) Consumer Characteristics

Cluster One is predominantly composed of adolescents aged 15–24 years, who are mostly students, unemployed, and have monthly expenditures ranging from IDR 500,000 to IDR 1,000,000. In contrast, Cluster Two is dominated by adults aged 35–54 years, who are employed and earn monthly incomes exceeding IDR 7,000,000, with monthly expenditures above IDR 3,000,000. Consumers in Cluster One tend to prefer processed products such as nuggets, whereas consumers in Cluster Two show a stronger preference for processed meatballs (bakso).

2) Purchase Intensity
Cluster One, which is dominated by students, consumes processed meat or poultry products more frequently than Cluster Two. Nevertheless, consumers in both clusters consistently consume processed meat products on a monthly basis.

3) Attitudes Toward Product Halalness (Halal Awareness)
The halal status of processed food products is considered a crucial and primary factor by consumers in both clusters, regardless of religious background (Muslim and non-Muslim). The presence of a halal certification label is also a major consideration when purchasing processed meat products.

4) Influence of External Parties
Family influence plays a significant role in shaping purchasing decisions among consumers in Cluster One, whereas consumers in Cluster Two are less influenced by family members. However, both clusters share a commonality in that religious leaders have a notable influence on their decisions to purchase processed meat products.

5) Sales Location
Consumers in both clusters find it easier to obtain processed meat products at supermarkets rather than traditional markets. This preference is particularly evident among consumers in Cluster Two, who feel more comfortable purchasing products at supermarkets compared to traditional markets.

6) Price and Promotion
Price is a significant factor influencing purchasing decisions in both clusters. However, promotional activities or discounts do not significantly affect consumers in either cluster. In particular, consumers in Cluster Two prioritize proximity of the sales location over promotional offers when deciding to purchase processed meat products.

7) Sensory Appeal
The visual appeal (e.g., color and appearance) of halal-certified processed poultry or meat products significantly influences purchasing decisions among consumers in both clusters.

8) Brand
The brand of processed meat products strongly influences purchasing decisions in both clusters. This similarity indicates that brand reputation and halal certification logos play an important role in shaping consumer trust and purchase intentions.

CONCLUSION

This research has applied the Fuzzy C-Means method to cluster consumers who purchase processed meat products. Two clusters were identified, with the highest PCI value of 0.9993799515177483 at $m = 2.4$. The error value was set at 0.0001 with 1000 iterations. The first cluster, dominated by teenagers, consumes processed meat products more frequently than the second cluster, which consists of adults. The influence of family on processed meat consumption is more dominant in the first cluster, while religious leaders have a significant impact on both clusters. The most consumed processed meat product in the first cluster is nuggets, while the second cluster prefers meatballs. The location of sales, brand, price, and sensory appeal of processed products significantly influence consumers in both clusters. The logo and halal certification of the product are the primary factors considered by consumers when shopping. However, promotions and discounts do not have a significant effect on consumer purchasing behavior. Consumers tend to prefer shopping at supermarkets rather than at traditional markets, particularly those close to their homes.

REFERENCES

[1] The Royal Islamic Strategic Studies Centre, "The Muslim 500: The World's 500 Most Influential Muslims," RISSC, 2024.

[2] BPJPH, "Tentang BPJPH," Kemenag RI, Jakarta, 2023.

[3] B. P. Statistik, "Produksi Daging Ayam Ras Pedaging menurut Provinsi (Ton), 2020-2022," BPS, 2023. [Online]. Available: https://www.bps.go.id/indicator/24/488/1/produksi-daging-ayam-ras-pedaging-menurut-provinsi.html.

[4] ARPI, "Cold Chain Data," 2022.

[5] D. Ghosh, "Comparative Analysis of K-Means and Fuzzy C-Means Algorithms," *IJACSA, vol. 4, no. 4,* 2013.

[6] A. ,. J. HR Mohammed, "Improved Fuzzy C-Mean Algorithm for Image Segmentation," *(IJARAI) International Journal of Advanced Research in Artificial Intelligence,* vol. 5, no. 6, 2016.

[7] G. J. X. L Hua, "A Novel Brain MRI Image Segmentation Method Using an Improved Multi-View Fuzzy C-Means Clustering Algorithm," *Frontiers in Neuroscience,* vol. 15, 2021.

[8] H. G. Deviyana Nurmin, "Penerapan Metode Fuzzy C-Means Pada Pengelompokan Kabupaten/Kota di Pulau Kalimantan Berdasarkan Indikator Kesejahteraan Rakyat Tahun 2020," *Jurnal EKSPONENSIAL,* vol. 13, no. 2, pp. 190-196, 2022..

[9] A. Kumar, "Bias-Corrected Intuitionistic Fuzzy C-Means With Spatial Neighborhood Information Approach for Human Brain MRI Image Segmentation," *IEEE Transactions on Fuzzy Systems,* vol. 30, no. 3, pp. 687-700, 2022.

[10] L. B. Y. W. Zhang, "Hybrid Missing Value Imputation Algorithms Using Fuzzy C-Means and Vaguely Quantified Rough Set,," *IEEE Transactions on Fuzzy Systems,* vol. 30, no. 5, pp. 1396 - 1408, 2022.

[11] Y. L. Yu, "Dynamic Background Subtraction Using Histograms Based on Fuzzy C-Means Clustering and Fuzzy Nearness Degree,," *IEEE Access,* vol. 7, pp. 14671 - 14679, 2019.

[12] P. Z. ,. Z. Roh, "Design of Iterative Fuzzy Radial Basis Function Neural Networks Based on Iterative Weighted Fuzzy C-Means Clustering and Weighted LSE Estimation,," *IEEE Transactions on Fuzzy Systems,,* vol. 30, no. 10, 2022.

[13] X. G. C. Wu, "A Novel Single Fuzzifier Interval Type-2 Fuzzy C-Means Clustering With Local Information for Land-Cover Segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* Vols. 14,, pp. 5903-5917, 2021.

[14] Siringoringo, "Peningkatan Performa Cluster Fuzzy C-Means PADA Pengklasteran Sentimen Menggunakan Particleswam Optimazion," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK),* vol. 6, no. 4, pp. 349-354, Juni 2019..

[15] L. F. E. Rouza, ""Implementasi Fuzzy C-Means Clustering dalam Pengelompokan UKM Di Kabupaten Rokan Hulu,," *Techno.COM,* vol. 19, no. 4, pp. 481-495, 2020.

[16] M. N. Sutoyo and A. T. Sumpala, "Penerapan Fuzzy C-Means untuk Deteksi Dini Kemampuan Penalaran Matematis," *Scientific Journal of Informatics,* vol. 2, no. 2, 2015.