East Asian Journal of Multidisciplinary Research (EAJMR)
Vol.2, No.4 2023: 1647-1660

Implementation of Fuzzy C-Means Algorithm with Optimized Parameter Grid for Clustering Electronic Product Sales

Rini Astuti[1*], Nining Rahaningsih[2], Umi Hayati[3], Cep Lukman Rohmat[4], Nana Suarna[5]
[1]Informatics Engineering, STMIK LIKMI Bandung, Indonesia
[2]Accounting Computerization, STMIK IKMI Cirebon, Indonesia
[3]Informatics Engineering, STMIK IKMI Cirebon, Indonesia
[4]Software Engineering, STMIK IKMI Cirebon, Indonesia
[5]Informatics Engineering, STMIK IKMI Cirebon, Indonesia
Corresponding Author: Rini Astuti riniastuti@likmi.ac.id

A R T I C L E I N F O

Keywords: Fuzzy C-Means, Implementation Clusters, Optimize Grid Parameters, Sales of Electronic Goods

Received : 05, February
Revised : 10, March
Accepted: 15, April

A B S T R A C T

The sales of electronic products have increased rapidly over the past few years. However, grouping products based on certain criteria is still an unresolved issue. Therefore, research is needed to develop more accurate clustering methods. Currently, the problem with electronic product clustering using the k-means method still has limitations, such as sensitivity to initial centroid values and inability to handle overlap between clusters. Therefore, research is needed to optimize the grid parameter of the Fuzzy C-Means algorithm to produce more accurate clustering. The purpose of this study is to implement the Fuzzy C-Means algorithm with optimized grid parameters to cluster electronic product sales more accurately. The method used in this study is an experimental research method. Electronic product sales data were obtained from specific stores, and the Fuzzy C-Means algorithm with optimized grid parameters was applied to cluster electronic products. The results show that implementing the Fuzzy C-Means algorithm with optimized grid parameters can produce more accurate electronic product clustering compared to the k-means method. By using optimized grid parameters, the Fuzzy C-Means algorithm can handle overlap between clusters and produce more stable centroids with a Dbi accuracy value of 0.510 for Numerical Measure and 0.611 for Mixed Measure.

DOI prefix: https://doi.org/10.55927/eajmr.v2i4.3929
ISSN-E: 2828-1519
https://journalformosapublisher.org/index.php/eajmr

INTRODUCTION

Electronic product sales have become an integral part of the modern industry today. In this digital era, customers have easy access to various electronic products such as smartphones, computers, tablets, and other electronic devices. In a store, there are many electronic products sold in large numbers, making it difficult for business managers to understand sales trends and customer behavior as a whole.

To overcome this problem, clustering can be used as a method for analyzing and understanding electronic product sales trends. Clustering is a data analysis method used to group similar data into homogeneous groups or clusters. In clustering, data is grouped into clusters that have similarities in certain characteristics, such as product attributes or features. The clustering method can help business managers understand customer behavior, sales trends, and consumption patterns.

However, conventional clustering algorithms often produce low performance, especially when applied to complex and high-dimensional data. Therefore, the Fuzzy C-Means (FCM) algorithm can be a better choice for clustering electronic product sales data. The FCM algorithm can overcome inconsistency and ambiguity in data and can produce more complex groups.

Previous research has been conducted to optimize clustering performance using the FCM algorithm. One related study is conducted by Zhang et al. (2019) entitled "An optimized FCM algorithm based on artificial bee colony and grey wolf optimization for image segmentation." This study used an optimized FCM algorithm with artificial bee colony and grey wolf optimization to perform image segmentation. The results showed that the optimized FCM algorithm produced better performance than the conventional FCM algorithm.

Furthermore, a study conducted by Jain et al. (2018) entitled "Fuzzy C-Means Clustering Based Segmentation of Brain MRI Images using Improved Fuzzy Entropy with Inverse Weightage Factor" also used the FCM algorithm for brain MRI image segmentation. This study developed a method to improve FCM algorithm performance using the inverse weightage factor of fuzzy entropy. The results showed that the proposed method could improve the accuracy of brain MRI image segmentation.

The proposed study will implement the FCM algorithm with optimized parameter grid for clustering electronic product sales. The parameter grid is a technique used to find optimal parameters in the FCM algorithm. By using this technique, more optimal and accurate clustering results will be obtained. The aim of this study is to analyze electronic product sales trends and customer behavior based on the clustering results.

This research will be conducted using electronic product sales data from an electronic store. The data will be analyzed using the FCM algorithm implemented with an optimized parameter grid. The clustering results will be analyzed to identify electronic product sales trends and customer behavior. In addition, a comparison of the clustering results between the FCM algorithm with an optimized parameter grid and the conventional FCM algorithm will be

conducted to show the advantages of implementing the FCM algorithm with an optimized parameter grid in clustering electronic product sales data.

The purpose of this study is to improve the performance of electronic product sales data clustering by using the FCM algorithm implemented with an optimized parameter grid. By using this technique, more accurate and optimal clustering results are expected to be obtained in analyzing electronic product sales trends and customer behavior. The description of electronic product sales is continuously increasing along with the advancement of technology. To manage business well, store managers need to understand sales trends and customer behavior as a whole. One method that can be used to analyze electronic product sales data is clustering. However, conventional clustering algorithms often produce low performance, especially when applied to complex and high-dimensional data.

METHODOLOGY

The research method that can be used in the study titled "Implementation of Fuzzy C-Means Algorithm with Optimized Parameter Grid for Clustering Electronic Product Sales" is as follows:

a. Literature Review Firstly, the researcher needs to conduct a literature review to understand the basic concept of clustering and fuzzy c-means algorithm, as well as its application in electronic product sales data. This will help the researcher in selecting the appropriate clustering technique to be used in this research.
b. Data Collection In this stage, it is necessary to collect electronic product sales data from reliable sources. This data will then be used to perform clustering using the fuzzy c-means algorithm.
c. Data Preprocessing After the data is collected, the researcher needs to preprocess the data to clean it from noise or outliers. This can be done using techniques such as filtering, smoothing, and so on.
d. Parameter Selection After the data is preprocessed, the researcher needs to select optimal parameters for the fuzzy c-means algorithm. This can be done using techniques such as grid search, where the researcher will try several parameter combinations to find the combination that provides the best clustering result.
e. Implementation of Fuzzy C-Means Algorithm This stage involves implementing the fuzzy c-means algorithm using programming languages such as Python or R. This algorithm will divide electronic product sales data into several different clusters based on their similarity characteristics.
f. Result Evaluation After clustering is performed, the researcher needs to evaluate the results using metrics such as silhouette score or sum squared error. This will help the researcher in evaluating the quality of the clustering produced and ensuring that the selected parameters are optimal.
g. Result Interpretation In this stage, the researcher needs to interpret the clustering results to gain insights into the pattern of electronic product

sales. The clustering results can be visualized in the form of graphs or plots to facilitate interpretation.[5]

The data analysis technique

Used to analyse the dataset of uninhabitable houses in data mining implementation involves the stages of the Knowledge Discovery in Databases (KDD) process[6], which consists of Data, Data Cleaning, Data Transformation, Data Mining, Pattern Evolution, and Knowledge, as shown in Figure 1.

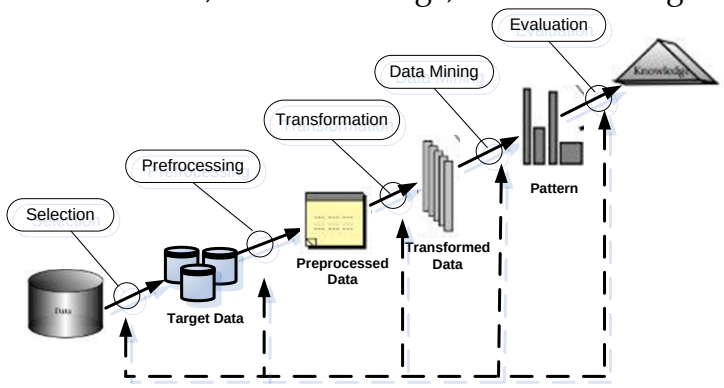

Figure 1. Data Analysis Techniques

The KDD stages from figure 1 are:

1. Data Selection this stage involves selecting relevant data for the research purpose. Relevant data may include electronic product sales data such as sales volume, price, and others.
2. Preprocessing this stage involves preprocessing data to clean data from noise or outliers, filling missing values, and selecting relevant features. Proper data preprocessing will improve the quality of the resulting clustering.
3. Transformation this stage involves transforming data to convert unstructured data into structured data. For example, converting qualitative data into quantitative data.
4. Data mining the data mining used in this research is clustering using the fuzzy c-means algorithm. Fuzzy c-means algorithm is a clustering technique that divides data into several clusters based on the similarity of characteristics between data.
5. Evaluation this stage involves evaluating the clustering results using metrics such as silhouette score or sum squared error. This will help researchers evaluate the quality of the resulting clustering and ensure that the selected parameters are optimal[7][8]

RESULTS

The research results will describe the process of clustering electronic product dataset through testing the Fuzzy C-Means Algorithm. This clustering process is carried out using machine learning through software such as RapidMiner Studio.

a. Data

The dataset used in this research consists of 900 records and 11 attributes of electronic products. The dataset was obtained from the Kaggle repository with the website https://www.kaggle.com/datasets/kiranbudati/mobile-prices-flipkart. The data was in the form of a soft file document in CSV format, as shown in table 1.

No	U_Id	Name	Offer_ Price	Original_ Price	Total_ Ratings	Total Reviews	RAM
1	22D33RGW	HPOMEN Ryzen7 Octacorea	99990	124283	0	0	16GBDD R5RAM
2	1X0V8DP0	Infinixx1 Seriescorei7 10thgen	46990	69999	128	17	16G BLP DDR4XRAM
3	EBK8ZBOF	ASUS Vivo Book15	33990	45990	3600	370	8G BDD R4RAM
4	2UWFCQ6Z	Asusvivo Book	43990	57990	2408	211	8GB DDR 4RAM
5	RHHI5DCG	ASUSTUF Gaming F15Corei/	47990	70990	1209	100	8GB DDR4RAM
6	T2LBXWSX	Hppavilion Ryzen5 Hexa Corea	55990	63539	8146	851	8GB DDR4RAM
7	RWIIUF8L	Hpcorei5 12thgen	58499	72331	301	27	16GB DDR4RAM
8	N0F1Q7EX	Infinixx1seriescorei7	46990	69999	128	17	16GBL PD DR4XRAM
9	D8P5OYHY	ASUSTUF Gaminga17 With90whr	51990	71990	350	47	8GB DDR4RAM
10	VR1DIKXD	Hpcorei 311t H Gen- (8GB/	40999	49508	1728	148	8GB DDR4RAM
..	…….	……..	……..	……..	…….	…….	……
900	JL6N2361	Infinixx1 Slimseries	46990	69999	80	20	16GBL P DDR4XRAM

Table 1: Electronic Dataset

b. Data Selection

To read the dataset in excel format, use the Read Excel operator as shown in Figure 2.

Figure 2. Excel Read Operators on Rapidminer

The parameters used for the Read Excel operator were the default parameters. From the result of the Read Excel operator, the following information was obtained.

No.	*Description*	*Informasi*
	Record	900
	Special Attribute	0
	Reguler Atribute	11
	Attribute :	
1.	u_id	polynominal
2	Name	polynominal
3	offer_price	integer
4.	original_price	integer
5.	total_reviews	integer
6	Rating	real
7.	item_link	polynominal
8.	created_at	date
9	processor	polynominal
10	Ram	polynominal
11	systemoperasi	polynominal

Table 2. Dataset statistics

To perform the first selection, which is to determine the ID in the dataset, the Set Role operator is used as shown in Figure 3.

Figure 3. Operator Set Role on Rapidminer

The parameters used in the Set Role operator are shown in the following table.

No.	**Parameter**	**Isi**
1.	Attribute name	u_id
2.	Target role	id

Table 3: Parameters and attributes selected in the Select Attribute operator

The next selection on the dataset is done using the Select Attributes operator, as seen in Figure 4.

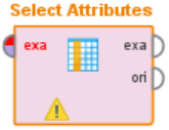

Figure 4. Attribute Select Operator on Rapidminer

The attributes that were not selected are:

1. Name, because this attribute is relevant to u_id
2. Item link, because this attribute is relevant to u_id.

The process model in Rapid Miner in the Selection step can be seen in figure 5.

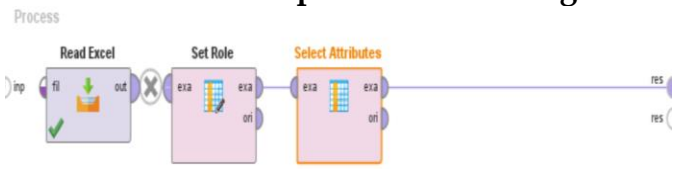

Figure 5. Model of the Selection step process in Rapid miner

c. Pre-processing

The data cleansing process or data cleaning for missing or inconsistent values is carried out in the pre-processing step. Before performing this process, an analysis is carried out to determine whether the selected dataset attributes have missing values or not and whether the data is consistent or not.[9]

From the result of the dataset statistics as shown in Figure 6, it is known that there are 3 attributes with missing values, namely the storage attribute (10), Office (519), and Warranty (1). To check the consistency of the dataset being used, it is directly examined per-record, and it shows that the dataset has consistent data values. It turns out that the Office attribute has an NA (no answer) value or in other words, it has no value or is missing.[10]

In the Pre-processing step, it is necessary to standardize or make consistent the "no answer" or "NA" values to blank values. This requires an operator called "Declare Missing Value" as shown in Figure 6.

Figure 6. Declare Missing Value operator in RapidMiner

The parameters used in the Declare Missing Value operator can be seen in table 3.

No.	Parameter	Isi
1.	Attribute filter type	All
2.	Nominal	Nominal
3.	Nominal value	NA

Table 3. Parameters in the Declare Missing Value operator.

Next, to turn these blank or missing values into a certain value, for example, a nominal value, we can use the Replace All Missing's operator as shown in Figure 7.

Figure 7. Replace All Missing's operator in Rapid Miner.

The process model in Rapid Miner for the Preprocessing step can be seen in Figure 8.

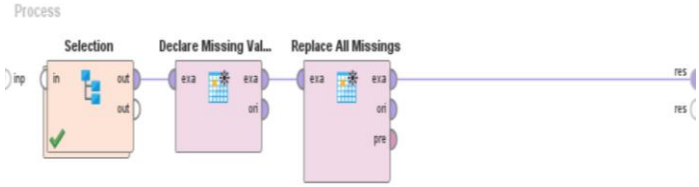

Figure 8. Preprocessing step process model in Rapid miner

d. Transformation

In the Transformation step, to convert data that is of polynomial or nominal type to numeric type, we can use the Nominal to Numerical operator as shown in Figure 9.

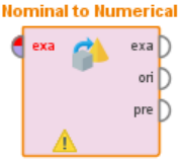

Figure 9. Nominal to Numerical operator in Rapid Miner

The parameters used in the Nominal to Numerical operator can be seen in Table 4

No.	Parameter	Isi
	Parameter	
1.	Attribute filter type	Subset
2.		Garansi
3.		Monitor
4.		Off_now
5.	Parametr	Office
6.	Coding type	Processor
7.		RAM
8.		SistemOperasi
9.		Storage

Table 4. Parameters and selected attributes in the Nominal to Numerical operator

The process model in Rapid Miner up to the Transformation step (Polynomial to Numerical) can be seen in Figure 10

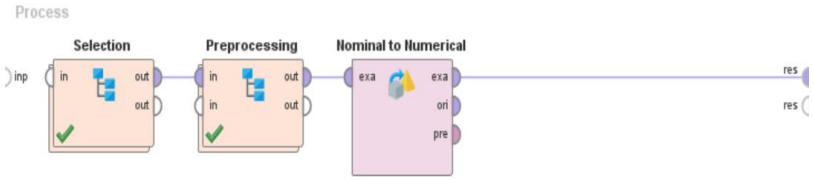

Figure 10. Process model up to the Transformation step

e. Data mining

In the data mining step, since we use Optimize Parameter (Grid), the Optimize Parameter (Grid) operator is used earlier before the Fuzzy C-Means operator

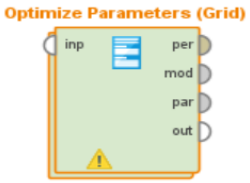

Figure 11. Optimize Parameter (Grid) operator on Rapid miner

Since the Optimize Parameter (Grid) operator is a type of sub process operator, the Fuzzy C-Means operator is installed inside this operator, as seen in Figure 12.

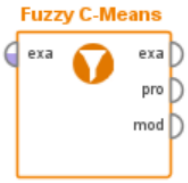

Figure 12.C-Means operator on Rapidminer

The parameters of the Fuzzy C-Means operator are ignored

After using the Fuzzy C-Means operator, use the Cluster Model from Data operator to create a Cluster Model from the existing cluster attribute. The Cluster Model from Data operator can be seen in Figure 13

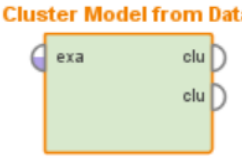

Figure 13. Cluster Model Operator from Data on Rapidminer

The parameters of the Cluster Model from Data operator are ignored. Next, the Cluster Distance Performance operator is also used. This operator is used to obtain the performance value of the parameters used in the Fuzzy C-Means operator and the resulting model. The Cluster Distance Performance operator is shown in Figure 14

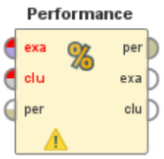

Figure 14. Cluster distance performance

Parameter in the Cluster Distance Performance operator is ignored. The process model in RapidMiner up to the Data Mining step is shown in Figure 15.

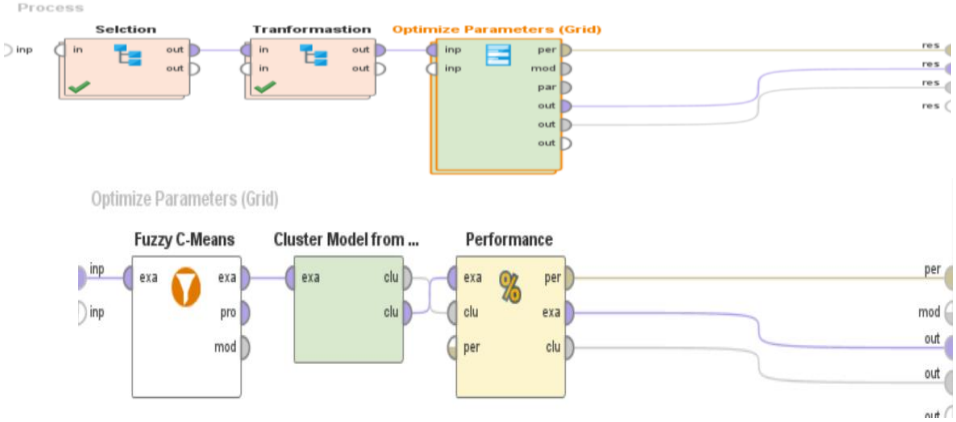

Figure 15. Process model up to the Data Mining step

The parameters used in the Optimize Parameter (Grid) operator are shown in the following table 5.

Parameter	Operator	Select Parameters	Value
Edit Parameter Setting	Fuzzy C-Means	Clusters	Min : 2 Max : 10 Steps : 10 Scale : Linear
		measure_type	NumericalMeasure MixedMeasure
	Performance (Cluster Distance Formance)	main_criterion	Davies Bouldin

Table 5:Selected parameters and attributes

f. Evaluation

The evaluation performed on the experimental results of the dataset obtained from the Numerical Measures resulted in the Davies Bouldin value as follows:

Literation	Clusters	Davies Bouldin
1	2	0.521
2	3	0.510
3	4	0.549
4	2	0.678
5	3	1.008

6	4	4.170
7	7	0.611
8	8	0.672
9	9	0.676
10	10	0.770

From these results it was concluded that:

a. The optimal number of clusters based on the Fuzzy C-Means algorithm is 3 clusters with Dbi = 0.510.
b. In measure type: Numerical Measure, Dbi = 0.510 and measure type: Mixed Measure, Dbi = 0.611. So the best measure type is Numerical Measure
    1. Cluster 0 = 608 item
    2. Cluster 1 = 202 item
    3. Cluster 2 = 90 item

Additional information on two graphic displays

1. Cluster vs Rating

Figure 16. Clusters vs Ratings

Information:

The members of cluster 0 have a better distribution than those of cluster 1 and 2, the distribution of members in cluster 1 and 2 are more abundant than cluster 1 and 2, while the rating of cluster 0 is better than cluster 1 and cluster 2.

2. Cluster vs Garansi

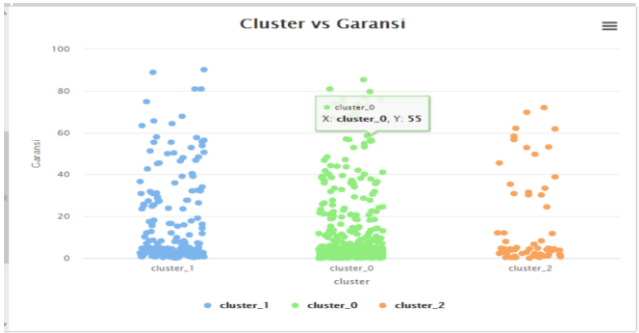

Figure 17. Clusters vs Guarantees

Information:
The distribution of group members in cluster 0 is better than cluster 1 and cluster 2, and the spread of group members in cluster 0 is greater than cluster 1 and cluster 2, while the rating of cluster 0 is better than cluster 1 and cluster 2. The better guarantee is in cluster 0 because its group spread is more compact.

CONCLUSIONS AND RECOMMENDATIONS

Based on the results of the electronic dataset clustering study using the Fuzzy C-Means algorithm, the following conclusions can be drawn:

1. The best information on the clustering of the Electronic dataset can be obtained based on the Davies Bouldin Index value generated from the Fuzzy C-Means algorithm, which is on measure type: Numerical Measure, Dbi = 0.510, and measure type: Mixed Measure, Dbi = 0.611.
2. The analysis and clustering of the Electronic dataset using the Fuzzy C-Means algorithm show that the best clustering results can be obtained with 3 clusters, where cluster 0 consists of 608 items, cluster 1 consists of 202 items, and cluster 2 consists of 90 items.
3. The number of members in each of the best clusters is as follows: Cluster 0 = 608 items, Cluster 1 = 202 items, Cluster 2 = 90 items.

REFERENCES

A. A. Setiawan1, "MULTICLASS SVM DENGAN OPTIMIZER PARAMETER GRID UNTUK MEMPREDIKSI PERFORMANCE STUDEN," IJIR, vol. 3, no. 1, pp. 36–41, 2022.

A. Dagoumas and G. C. Christoforidis, "An Improved Fuzzy C-Means Algorithm for the Implementation of Demand Side Management Measures," energgies MDPI, vol. 10, no. September, 2017, doi: 10.3390/en10091407.

G. Nabila, S. Putri, D. Ispriyanti, T. Widiharih, D. Statistika, and U. Diponegoro, "IMPLEMENTASI ALGORITMA FUZZY C-MEANS DAN FUZZY POSSIBILISTICS C-MEANS UNTUK KLASTERISASI DATA TWEETS PADA AKUN TWITTER TOKOPEDIA," J. GAUSSIAN, vol. 11, pp. 86–98, 2022.

J. Chen, H. Zhang, D. Pi, M. Kantardzic, Q. Yin, and X. Liu, "A Weight Possibilistic Fuzzy C-Means Clustering Algorithm," Hindaw Res. Artic. A, vol. 2021, 2021.

J. Yin, H. Chang, D. Wang, H. Li, and A. Yin, "Fuzzy C -Means Clustering Algorithm-Based Magnetic Resonance Imaging Image Segmentation for Analyzing the Effect of Edaravone on the Vascular Endothelial Function in Patients with Acute Cerebral Infarction," vol. 2021, 2021.

L. Rahmadhani, A. Djunaidy, and A. Mukhlason, "Evaluasi Kinerja Pemasok Menggunakan Fuzzy C-Means Clustering dan AHP di CV Delta Raya," J. Tek. ITS, vol. 10, no. 2, 2021.

M. Priyono, T. Sulistyanto, K. Suharsono, and D. A. Nugraha, "Monitoring dan Kendali Peralatan Elektronik Menggunakan Logika Fuzzy Melalui Website Dengan Protokol HTTP," J. SMARTICS, vol. 2, no. 2, pp. 49–54, 2018.

R. Siringoringo, "PENINGKATAN PERFORMA CLUSTER FUZZY C-MEANS PADA PENGKLASTERAN SENTIMEN MENGGUNAKAN PARTICLE AN IMPROVED FUZZY C-MEANS FOR SENTIMENT CLUSTERING BASED ON PARTICLE SWARM OPTIMIZATION," J. Teknol. Inf. dan Ilmu Komput., vol. 6, no. 4, pp. 349–354, 2019, doi: 10.25126/jtiik.2018561090.

T. X. Pham, P. Siarry, and H. Oulhadj, "Image Clustering Using Improved Particle Swarm Optimization," pp. 359–373, 2018.

Y. Nugraheni, "DATA MINING USING FUZZY METHOD FOR CUSTOMER RELATIONSHIP MANAGEMENT," LONTAR Komput., vol. 4, no. 1, pp. 188–200, 2013.