Handling Noise Data with PCA Method and Optimization Using Hybrid Fuzzy C-Means and Genetic Algorithm

Risa Widianti[1], Sugiyarto Surono[2*], Kais Ismail Ibraheem[3]

[1,2]*Mathematics Departement, Ahmad Dahlan University, Indonesia*

[3]*Computers Science of Department, Mosul University, Iraq*

*corr_author: sugiyarto@math.uad.ac.id

Abstract - The significance of machine learning (ML) and data mining techniques particularly clustering is examined in this research, in managing large data sets for customer segmentation in the retail sector. The research emphasizes the challenges posed by data noise and proposes a solution using Principal Component Analysis (PCA) to improve accuracy. This study introduces a hybrid approach that combines Fuzzy C-Means (FCM) with genetic algorithms for optimization in customer segmentation, and suggests further research on the optimal number of clusters and data noise elimination. By addressing data noise, the proposed PCA-based method achieved a higher accuracy rate of 98% compared to 93% without PCA. This finding underscores the effectiveness of PCA in noise reduction, improving clustering accuracy. This research contributes to the advancement of customer-focused business strategies through better data analysis and interpretation. The proposed approach has potential applications in areas including data analysis, pattern recognition, and image processing, highlighting its relevance in the contemporary business environment.

Keywords: noise data, Principal Component Analysis, Fuzzy C-Means, Genetic Algorithm, customer segmentation

I. INTRODUCTION

In the era of globalization and intense business competition, a deep understanding of customers is key to maintaining a competitive advantage. Digital marketing is growing rapidly in the retail sector, utilizing big data to meet customer needs [1]. Big data refers to data that is too big, fast, and complex to manage. Organizations today collect data from a variety of sources, including transactions, social media, and more. Advances in computing and mathematics have enabled effective analysis of big data. Client data is considered a strategic asset with a focus on responsible use for economic value as well as maximum efforts to maintain security [2].

Currently, data mining is usually widely used in companies that have a large number of customers, such as financial companies, telecommunications companies, and marketing organizations [3]. Data mining is becoming one of the powerful new technologies that have developed. This technology helps individual users and companies get data from large data sets [4]. One of the common problems that affects data quality is noise. However, the ever-increasing scale of data is making challenges more difficult for conventional solutions designed to deal with noise [5]. Data noise in data mining is irrelevant information in a dataset that can interfere with the accuracy of prediction and analysis, including class noise in target variables and attribute noise in independent variables. Identifying and handling noise in datasets is crucial to ensuring optimal classification accuracy and prediction results [6].

Reducing the highest value dimensional data set to a low dimensional data set with filters or removing redundant information and noise data is a method for solving this problem known as dimensionality reduction [7]. The most frequently used approach for dimensionality reduction is Principal Component Analysis (PCA). Thus, data becomes easier to interpret and processed faster. PCA is a statistical technique that preserves as much information as possible while removing noise. This technique reduces the number of high dimensions in large data sets to fewer, thus speeding up the storage and processing process [8]. Although PCA is effective in reducing dimensionality and removing noise, other issues such as sensitivity to noise in the context of data clustering still need to be addressed. Fuzzy C-Means (FCM) is a significant clustering algorithm with a wide range of applications including retail market data analysis, network monitoring, web usage mining, and stock market prediction [9]. Apart from using data grouping, for handling noise it is also necessary to use optimization methods to improve

performance and efficiency in solving complex problems. the use of optimization methods such as Genetic Algorithm (GA) has proven effective in overcoming these challenges. GA enriches this approach with its ability to perform probabilistic searches in a wide search space, strengthening FCM's ability to handle noise-sensitive data clustering problems. However, to overcome these challenges, the use of optimization methods such as GA has been proven to be an effective global optimization approach due to its ability to perform probabilistic searches in a wide search space [10].

In [11], Researchers used a hybrid method to estimate semiconductor manufacturing cycle time, integrating GA and FCM methods. The FCM - GA method is a new unsupervised soft classification technique that uses GA to improve the classification capabilities of FCM. This research also recommends determining and deleting invalid records due to data noise because it can affect accuracy results.

Based on research [11], research will be carried out that focuses on handling noise data using the PCA method, this method is to improve data quality by reducing or eliminating irrelevant information. After handling noise from the data, optimization will be carried out using the FCM-GA hybrid method. which can help solve problems with data and improve prediction accuracy.

II. METHOD

This In this research, we will try to handle data noise with the aim of improving the quality of data analysis and interpretation by reducing or eliminating irrelevant or unwanted information using the PCA method. After that, applying the hybrid FCM algorithm and GA to find clusters, especially in customer segmentation, can help overcome the challenges in the data after cleaning up the noise and also improve the prediction accuracy. Fig. 1 shows the specifics of the steps involved in the research.

A. Dataset

This research uses datasets taken from Kaggle about customer segmentation. The dataset is based on customer segmentation variables, which are used as input variables in this research.

B. Preprocessing Data

Data normalization is used in the data preprocessing step where the process of converting the numerical value of a feature in the dataset to a more common value is usually performed at the data preprocessing stage. This procedure results in a range of attribute values [0, 1], which is intended to shorten the training time of the algorithm and improve the stability of the final model [12]. The min-max normalization used in this study is (1) as follows [13].

$$Q^* = \frac{Q - min\,(Q)}{max(Q) - min(Q)} \quad (1)$$

where,
Q^* = Data normalization results
Q = Original data
$\min(Q)$ = Minimum value of Q
$\min(Q)$ = Maximum value of Q

C. Principal Component Analysis

Karl Pearson first presented PCA in the early 1900s. A collection of connected variables is reduced to a smaller set known as principle components (PCs) using PCA [14]. PCA works to reduce the input data to facilitate and speed up the classification process [15]. PCA uses the linear combination of the initial variables to compute new variables known as principle components in order to reach its goal [16].

Stages of the PCA process:

- Mean calculation
- Calculation of the Covariance Matrix
- Calculation of the Eigenvector and Eigenvalue of the Covariance Matrix
- Principal Component and Feature Vector Analysis and Selection

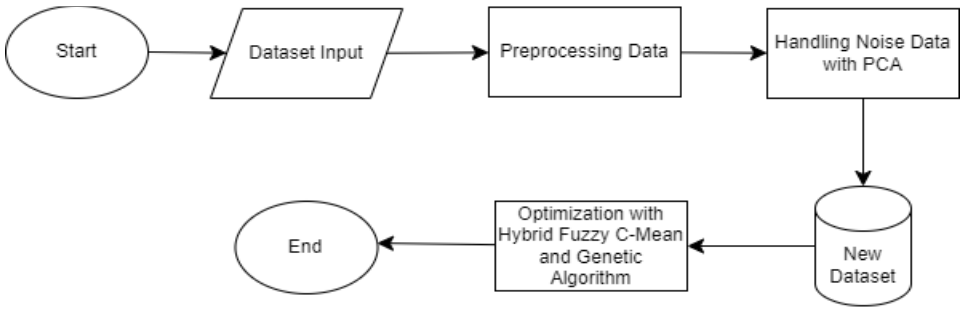

Fig. 1 Research flow

D. Fuzzy C-Means

FCM a soft partitioning technique, is used to classify tasks into appropriate groups based on similar attribute values. It's used to put patterns or data into discrete clusters, letting a single data point be a part of multiple clusters with different membership values. The cluster center is updated using this membership value, which also shows the distance between each point inside each cluster. In this FCM clustering approach, There is a fixed number of clusters, and one or more of them are assigned to each data point. Due to the fact that the FCM technique relies on minimizing the c-means function, an objective function, Consider it a fuzzified variant of the k-means clustering algorithm. The three required input parameters are the halting tolerance $\phi > 1$, the fuzziness exponent value $m > 1$ and the number of clusters G [11].

Fuzzy C-Means algorithm as follows:

1) Minimizing FCM c-means function: The following illustrates how to minimize the c-means function, an objective function: The number of fuzzy similarities fozr job records $\mathcal{R}$ across all clusters is determined by the goal function J. Keep in mind that FCM modifies the centers of every cluster at every iteration. The goal function J is displayed by (2) with the euclidean distance is d_{ic}.

$$J = \sum_{c \in G} \sum_{i \in \mathcal{R}} u_{ic}^{m} \cdot d_{ic} \quad (2)$$

2) For every task record: update the membership value u_{ic} towards i in cluster c. It should be noted that the procedure initializes the membership value u_{ic} at random so that $u_{ic} > 0$ and $\sum_{c \in G} u_{ic} = 1$. The necessary clustering tolerance limit is indicated by the fuzziness coefficient $1 < m < \infty$. Updates to the membership value u_{ic} are made by (3)

$$u_{ic} = \frac{1}{\sum_{I \in G} \left(\frac{d_{ic}}{d_{Ic}}\right)^{\frac{2}{m-1}}} \quad (3)$$

3) Identifying the cluster centers: Using (4), get the center dimension K for every cluster. Then, use (5) to update the membership value u_{ic} and each cluster's Euclidean distance from the cluster center c.

$$z_{cb} = \frac{\sum_{i \in \mathcal{R}} u_{ic}^{m} \cdot v_{ib}}{\sum_{i \in \mathcal{R}} u_{ic}^{m}} \quad (4)$$

$$d_{ic} = \sqrt{\sum_{b \in \mathcal{B}} [W_b \cdot (v_{ib} - z_{cb})]^2} \quad (5)$$

It should be mentioned that each property has a distinct value variation, and the most significant attribute does not exceed the smaller attributes. It is suggested that all attributes in the original data be placed into the same range. This prediction error decrease, w_b it can be calculated by applying the following (6).

$$w_b = \frac{100}{\max(v_{\mathcal{R}b}) - \min(v_{\mathcal{R}b})} \quad (6)$$

4) Termination condition: A gradual adjustment to the objective function J value, which controls whether the loop will continue or not. When ϕ designates the FCM termination tolerance, the FCM will terminate if

$$|J^{s+1} - J^{s}| < \phi$$

E. Hybrid Fuzzy C-Means and Genetic Algorithm Design

Preventing cluster centers from rapidly converging into local extrema can improve the performance of FCM, GA has been added. Numerous prior investigations have demonstrated the clear benefits of the combined GA and FCM approach. Here, we create a GA method that combines FCM with GA, based on FCM. We employ the following strategies: adaptive crossover, nonlinear rank selection, mutation selection, and center-based string coding (Fig. 2).

1) Chromosome structure: Selecting a useful chromosomal structure is crucial for solving GA.

2) Fitness function: Until the fitness value converges or the maximum number of generations is reached, individuals (chromosomes) are assessed and evaluated using the fitness function. As in (7)

$$F = \frac{10^6}{\sum_{c \in G} \sum_{i \in \mathcal{R}} {u_{ic}}^{m} \cdot d_{ic}} \quad (7)$$

3) Genetic operator: The following genetic operators are what motivate this search procedure: The following genetic operators are what motivate this search procedure:

- Selection operator: To choose which people will receive genetic surgery, a constant ratio selection procedure is employed.
- Crossover operator: We start by creating random numbers between 0 and 1, then we compare the generated numbers to the crossover probability P_c. The crossover operator is used to two randomly selected parents in order to produce two more children for P_c.
- Mutation operator: We produce a random value for each individual in the interval [0, 1] and compare it with the mutation probability P_m. If $< P_m$, we use single point mutation to change the individual by substituting a new random K-dimensional center gene for the randomly chosen gene.

4) The following steps should be applied to determine the Hybrid Fuzzy C-Means and Genetic Algorithm :

- Set the values for the following parameters: fuzziness value m, number of clusters G, number of generations GN, population size PS, crossover probability P_c, mutation probability P_m, and termination tolerance ε.
- Population initialization. Utilize the FCM clustering technique to create chromosomes, as many as PS. To generate chromosomes, using $\gamma_{ic}, 0 \leq \gamma_{ic} \leq 1$ we will determine the membership values for every cluster as in (8).

$$u_{ic} = \frac{\gamma_{ic}}{\sum_{c \in G} \gamma_{ic}} \quad (8)$$

- Making use of genetic operations. After that, each person's fitness value function F (Equation 9) can be determined. Next, in order to increase population diversity, we apply genetic operators, specifically crossover, mutation, and selection operators.
- Appliying optimal preservation. Fitness values will be revised in every generation following genetic surgery on each individual. Higher fitness value individuals will be chosen for survival.
- Check the termination condition. In this study, the iteration will end when the fitness variance value is reached or after a predetermined number of GN generations. Evolution will come to an end if either of these circumstances is satisfied. If not, we go back to step 3. The variance ε for a population PS, denoted by p, is computed using (9).

$$\varepsilon = \sum_{p=1}^{PS} \left(F_P - \frac{\sum_{p=1}^{PS} F_p}{PS} \right)^2 \quad (9)$$

We may obtain each cluster's K-dimensional center and membership value once the evolution procedure is finished. Next, if a variable's membership value is greater than a predetermined threshold value δ, associate it with every cluster. For every cluster, an extra binary variable called ϵ_{ic} is employed to determine the approximate quality of customer segmentation. One can calculate c for each cluster in the following way (10) and (11).

$$et_c = \frac{\sum_{i \in \mathcal{R}} u_{ic} \cdot ct_i \cdot \epsilon_{ic}}{\sum_{i \in \mathcal{R}} u_{ic} \cdot \epsilon_{ic}} \quad (10)$$

$$where\ \epsilon_{ic} = \begin{cases} 1 & if\ u_{ic} > \delta \\ 0 & otherwise \end{cases} \quad (11)$$

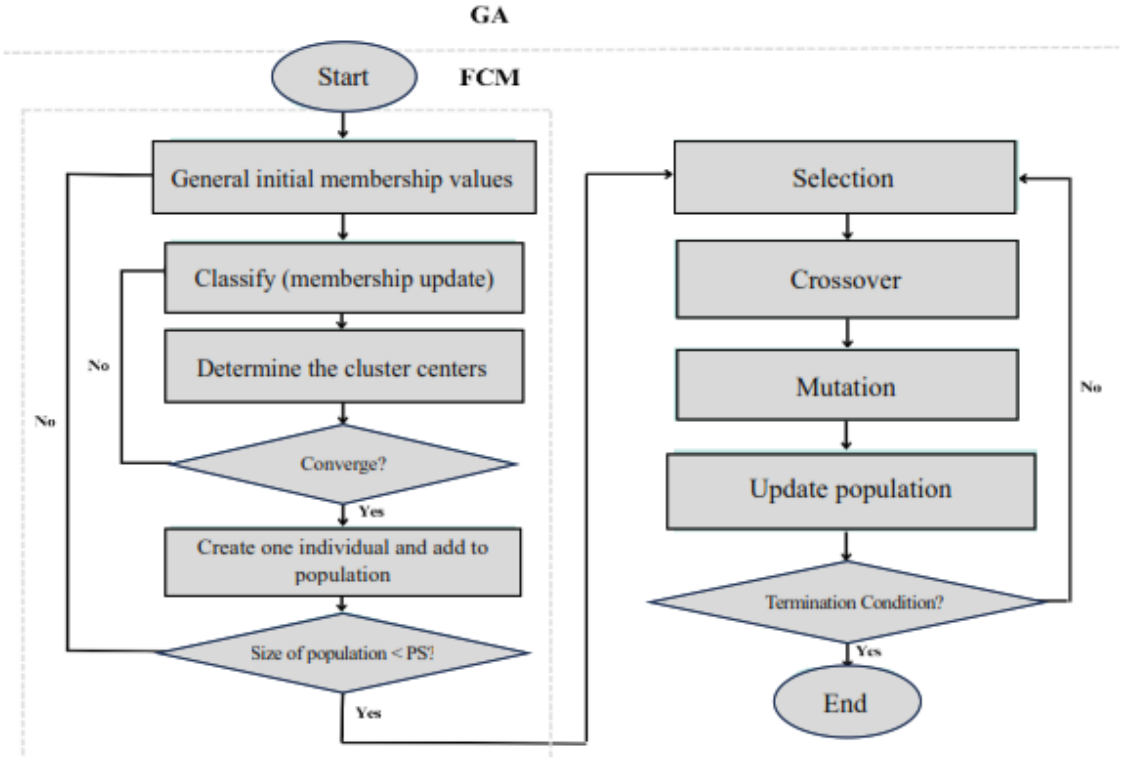

Fig. 2 Flowchart of FCM-GA

III. RESULT AND DISCUSSION

In this part of the research, data preprocessing was previously carried out first, which aims to prepare raw data so that it is ready for further processing by algorithms or data analysis models. Some reasons why data preprocessing is very important before processing data are as follows:

A. Handling missing values

Raw data may also usually contain missing or empty values. Data preprocessing makes it possible to handle missing values by filling, deleting, or swapping them out so the data is prepared for the analysis model to process.

B. Data standardization or normalization

Data analysis algorithms require data that has been standardized or normalized to have a uniform scale so as to provide optimal results. Preprocessing makes it possible to perform this standardization or normalization.

In Fig. 3, a dataset consisting of 18 variables, there are many outliers that can interfere with prediction accuracy. These outliers can affect the distribution of the data, resulting in inaccurate and biased estimates in the prediction model. The presence of outliers can cause statistical and machine learning models to give inconsistent and unreliable results. Therefore, it is important to conduct an in-depth analysis of outliers and consider strategies to manage them. Table I shows the Eigen value, total variance, and cumulative.

Based on Table I, no eigenvalue is greater than or equal to one. Therefore, the principal component is determined using the PKV value in the following calculation

$$PKV = \frac{\sum_{n=1}^{5} \lambda_n}{\lambda_1 + \lambda_2 + \cdots + \lambda_{16}} \times 100\%;\ n = 1,2,3,4,5$$

$$= \frac{0.3289 + 0.0932 + 0.0851 + 0.0509 + 0.0461}{0.6625} \times 100\%$$

$$= \frac{0.6042}{0.6625} \times 100\ \% = 0.9113$$

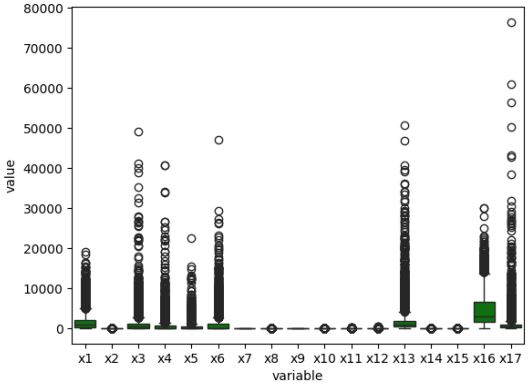

Fig. 3 Outlier data

TABLE I
EIGEN VALUES, TOTAL VARIANCE, AND CUMULATIVE

Component	Eigenvalues, Total Variance, and Cumulative		
	Eigen-values (λ)	Total Variance (%)	Cumulative (%)
0	0.3289	49.5972	49.5972
1	0.0932	14.0540	63.6512
2	0.0851	12.8444	76.4957
3	0.0509	7.6765	84.1723
4	0.0461	6.9634	91.1357
5	0.0238	3.5885	94.7243
6	0.0120	1.8193	96.5436
7	0.0089	1.3528	97.8964
8	0.0043	0.6554	98.5519
9	0.0036	0.5509	99.1028
10	0.0018	0.2841	99.3870
11	0.0012	0.1858	99.5728
12	0.0000	0.0000	99.5728
13	0.0004	0.0720	99.6449
14	0.0007	0.1057	99.7506
15	0.0008	0.1225	99.8732
16	0.0008	0.1267	100.0000

From the PKV calculation, a value of 91.13% was obtained so that it can be seen that the main components obtained in this study were 5 components. These five components are shown based on the eigenvector and are referred to as the loading value / coefficient on the main component formed. The amount of variance of a variable that can be explained by the main component is called loading. A large variance can be interpreted as having a large influence on the main component. Loading of the 5 main components is shown in Table II.

Table III presents the results of the cluster centers determined based on the data that has been processed through the PCA method into 4 clusters. This process is important because PCA helps to reduce the most significant dimensions of the data so that the resulting cluster centers are more representative and accurate in capturing the underlying structure of the dataset. Thus, the PCA results provide a solid basis for determining the starting point of the clustering algorithm so as to improve the effectiveness and efficiency of the overall data clustering process.

Table IV presents the results of the PCA method. The method successfully reduced the dimensionality of the customer segmentation data from the original 17 components to 5 principal components that cover most of the variation in the original data. This process

condenses the information, making it easier to analyze without significant loss of meaning. By reducing the dimensionality, PCA helps in dealing with noise in the data, removing less relevant or redundant features that can lead to overfitting and difficult interpretation. The results of this method enable a better understanding of the main structures in customer segmentation data, facilitating more accurate and efficient customer grouping based on the main patterns identified by PCA.

Table V shows the values used in this study. These values are used to set parameters for the optimization method, namely hybrid FCM-GA in an effort to classify data, perform optimization, or modeling with the aim of achieving optimal and convergent results.

TABLE II
LOADING / MAIN COMPONENT COEFFICIENT

No.	Loading /Main Component Coefficient				
	PC_1	PC_2	PC_3	PC_4	PC_5
0	-0.0150	-0.0987	0.1611	0.0537	-0.0971
1	0.1017	-0.2662	0.3741	0.4212	-0.7061
2	0.0328	-0.0476	-0.0100	0.0010	0.0087
3	0.0199	-0.0584	-0.0150	-0.0047	0.0096
4	0.0353	0.0022	0.0052	0.0108	0.0016
5	-0.0168	-0.0133	0.0268	0.0050	-0.0488
6	0.6855	-0.0805	0.0472	-0.1473	0.0571
7	0.2426	-0.8227	-0.1687	-0.1926	0.1454
8	0.6371	0.4232	0.2259	0.0138	0.0239
9	-0.0740	-0.0559	0.1236	-0.0050	-0.2315
10	-0.0203	-0.0193	0.0398	0.0097	-0.0681
11	0.0748	-0.0636	0.0088	-0.0022	0.0225
12	0.0138	-0.0431	-0.0085	0.0266	-0.0054
13	0.1940	0.1127	-0.8495	0.3326	-0.3164
14	0.0369	-0.1080	0.1088	0.8027	0.5526
15	0.0296	-0.1307	-0.1307	0.0673	0.0064
16	0.0000	-0.0030	0.0215	0.0081	-0.0052

TABLE III
PCA CLUSTER CENTER WITH HYBRID FCM-GA

Cluster	PC_1	PC_2	PC_3	PC_4	PC_5
1	0.7696	-0.4090	0.3954	1.0701	-0.1224
2	0.7762	-0.4110	0.3914	1.0677	-0.1202
3	0.7745	-0.4029	0.3935	1.0706	-0.1205
4	0.7647	-0.4104	0.3977	1.0694	-0.1220

TABLE IV
PRINCIPAL COMPONENT (PC) RESULT

No	PC_1	PC_2	PC_3	PC_4	PC_5
0	0.2891	-0.3089	0.4422	1.1264	-0.0135
1	0.1627	-0.3872	0.3136	1.2867	-0.2225
2	1.0758	-1.3275	0.3837	0.9081	0.0389
3	0.1826	-0.4010	0.3589	1.0637	0.1002
⋮	⋮	⋮	⋮	⋮	⋮
8946	1.3211	0.0001	0.6103	0.2880	-0.6286
8947	1.1316	0.0160	0.2898	0.3229	-0.6036
8948	0.1252	-0.2023	0.1140	0.4349	-0.6944
8949	0.6767	-0.8062	0.1999	0.0567	-0.3882

TABLE V
RESEARCH PARAMETERS

Parameters	Value
Fuzziness exponent (m)	2
Number of Cluster (G)	4

Table VI shows the training testing data, accuracy, Precision, and Recall results. Training data is used to build a model and get the appropriate weights. While the testing data is used to determine the accuracy of the results with the actual value. Accuracy is the level of closeness between the value obtained and the actual value. Precision is the match between the part of the data taken and the information needed. Recall is the success rate of the system in finding back information. It can be seen the difference in accuracy between using the PCA method and not. Using the PCA method produces high accuracy because it is used to handle data noise. The data noise greatly affects the accuracy results.

TABLE VI
ACCURACY CALCULATION

Method	Partition	Accuracy	Precision	Recall
	training : testing			
Hybrid FCM-GA	80 : 20	93%	94%	93%
PCA and Hybrid FCM-GA	80 : 20	98%	97%	97%

IV. CONCLUSION

The "Credit Card Dataset for Clustering" is a consumer segmentation dataset that was obtained from Kaggle and used in this study. Approximately 9,000 active credit card customers' usage patterns over the previous six months are summed together in the dataset, which includes 18 behavioral factors at the customer level. Table 1 shows the client information. In this paper, an approach using the PCA method is proposed to deal with data noise problems so as to provide the best accuracy results. The consumer segmentation dataset provided the information needed in this study. But overall, the dataset that has been handled by noise data can produce good accuracy. We can see in Table 5 the comparison of results between using the PCA method and not using the PCA method before clustering with the hybrid fuzzy C-means method and genetic algorithm. The resulting accuracy is 98% for those using the PCA method, while for those not using the PCA method, it is 93%. Considering the comparison's accuracy findings, the PCA method can help in handling data noise so that the resulting accuracy is good. As for clustering using Hybrid FCM, the data used previously still has data noise, so it produces less than maximum accuracy.

ACKNOWLEDGEMENT

The researcher would like to thank all those who have supported and participated in this research. Thank you for your advice, criticism, and input during the research.

REFERENCES

[1] J. Zhou, J. Wei, and B. Xu, "Customer segmentation by web content mining," *J. Retail. Consum. Serv.*, vol. 61, no. March, p. 102588, 2021, doi: 10.1016/j.jretconser.2021.102588.

[2] J.-M. Sahut, D. Schweizer, and M. Peris-Ortiz, "Technological Innovations to Ensure Confidence in the Digital World," *SSRN Electron. J.*, no. February 2023, 2022, doi: 10.2139/ssrn.4160924.

[3] A. Hadi, "Segmentasi Pelanggan Internet Service Provider (ISP) Berbasis Pillar K-Means," *J. Ilm. Teknol. Inf. Asia*, vol. 13, no. 2, p. 151, 2019, doi: 10.32815/jitika.v13i2.413.

[4] M. A. Jassim and S. N. Abdulwahid, "Data Mining preparation: Process, Techniques and Major Issues in Data Analysis," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1090, no. 1, p. 012053, 2021, doi: 10.1088/1757-899x/1090/1/012053.

[5] D. García-Gil, J. Luengo, S. García, and F. Herrera, "Enabling Smart Data: Noise filtering in Big Data classification," *Inf. Sci. (Ny).*, vol. 479, no. 2019, pp. 135–152, 2019, doi: 10.1016/j.ins.2018.12.002.

[6] S. Gupta and A. Gupta, "Dealing with noise problem in machine learning data-sets: A systematic review," *Procedia Comput. Sci.*, vol. 161, pp. 466–474, 2019, doi: 10.1016/j.procs.2019.11.146.

[7] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction," *J. Appl. Sci. Technol. Trends*, vol. 1, no. 2, pp. 56–70, 2020, doi: 10.38094/jastt1224.

[8] B. M. S. Hasan and A. M. Abdulazeez, "A Review of Principal Component Analysis Algorithm for Dimensionality Reduction," *J. Soft Comput. Data Min.*, vol. 2, no. 1, pp. 20–30, 2021, doi: 10.30880/jscdm.2021.02.01.003.

[9] J. Chen, H. Zhang, D. Pi, M. Kantardzic, Q. Yin, and X. Liu, "A Weight Possibilistic Fuzzy C-Means Clustering Algorithm," *Sci. Program.*, vol. 2021, 2021, doi: 10.1155/2021/9965813.

[10] K. do Prado Ribeiro, C. H. Fontes, and G. J. A. de Melo, "Genetic algorithm-based fuzzy clustering applied to multivariate time series," *Evol. Intell.*, vol. 14, no. 4, pp. 1547–1563, 2021, doi: 10.1007/s12065-020-00422-8.

[11] G. M. Lee and X. Gao, "A hybrid approach combining fuzzy c-means-based genetic algorithm and machine learning for predicting job cycle times for semiconductor manufacturing," *Appl. Sci.*, vol. 11, no. 16, 2021, doi: 10.3390/app11167428.

[12] Ibnu Daqiqil Id, *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python*. UR PRESS Riau, Indonesia, 2021.

[13] J. Jiawei, H., Kamber, M., & Pei, *Data Mining Concepts and Techniques Third Edition*. 2012.

[14] N. H. Timm, *Applied Multivariate Analysis*. 2002.

[15] W. S. Hasibuan, "Penerapan Metode Fisherface Untuk Mendeteksi Wajah Pada Citra Pasfoto," *Bull. Electr. Electron. Eng.*, vol. 1, no. 3, pp. 122–126, 2021.

[16] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 4, pp. 433–459, 2010, doi: 10.1002/wics.101.