# 254-99Z_Article Text-1053-1-6-20210301.docx

# COMPARISON OF FUZZY SUBTRACTIVE CLUSTERING DAN FUZZY C-MEANS

**ᵃAnnisa Eka Haryati, ᵇ Sugiyarto, ᶜRizki Desia Arindra Putri**

ᵃ Department of Masters in Mathematics Education, Universitas Ahmad Dahlan, Indonesia
ᵇ,ᶜ Department of Mathematics, Universitas Ahmad Dahlan, Indonesia
E-mail: sugiyarto@math.uad.ac.id

## Abstract

*Multivariate statistics have related problems with large data dimensions. One method that can be used is principal component analysis (PCA). Principal component analysis (PCA) is a technique used to reduce data dimensions consisting of several dependent variables while maintaining variance in the data. PCA can be used to stabilize measurements in statistical analysis, one of which is cluster analysis. Fuzzy clustering is a method of grouping based on membership values that includes fuzzy sets as a weighting basis for grouping. In this study, the fuzzy clustering method used is Fuzzy Subtractive Clustering (FSC) and Fuzzy C-Means (FCM) with a combination of the Minkowski Chebysev distance. The purpose of this study was to compare the cluster results obtained from the FSC and FCM using the DBI validity index. The results of the cluster evaluation obtained were 3.9164 for FSC and 1.6468 for FCM. This shows that the results of clustering using FCM are better than FSC.*

*Key words: clustering, fuzzy clustering, fuzzy c-means, fuzzy subtractive clustering, principal component analysis*

## INTRODUCTION

A problem which occur in multivariate statistic is analyzing huge data consisting a lot of variables to acquire fewer variables (dimension reduction). However, the process should maintain some information which already included in the data itself. One of the popular method to apply those dimension reduction in multivariate statistic is Principal Component Analysis (PCA) [1]. PCA is a method to reduce the dimension consisting of a few dependent variables without transforming its data variants [2]. This method can be applied by transforming the data into new variables (main component) which not correlated to maximum variants. Based on Jolliffe[3], PCA method can be used to stabilize the statistics analysis measurement such as regression analysis or cluster analysis.

Clustering is a classification process applied to certain set of data which its cluster label are still unknown[4]. Clustering method will classify certain data into the same cluster, where each of the data have some certain level of similarity with the others. Hence, good cluster will have high degree of data similarity in its cluster and high degree of difference in another cluster [5].

Fuzzy clustering is one of the clustering method where its classification process is based on membership degree and includes set of fuzzy as its measurement base. Every data will have the probability to be included as a member of any cluster. This means that each of the data will not absolutely induced in only one cluster[6]. There are many technique which can be applied to do fuzzy clustering, such as Fuzzy Subtractive Clustering and Fuzzy C-Means.

Based on Kusumadewi and Purnomo[7] Fuzzy Subtractive Clustering (FSC) is one of the clustering method which applied on certain set of data where the number of cluster to be formed through this process are still unknown. The basic concept of FSC is by determining the coordinates for each data which have the highest

density. Coordinate with the highest number of neighboring data will be chose as the cluster centroid. Furthermore, the density level of the coordinate which chosen as the cluster centroid will be reduced and the algorithm will select another coordinates which have the most neighboring data to be chosen as the new cluster centroid. Fuzzy C-Means (FCM) is a fuzzy based clustering method where the number of cluster which will be formed are initially determined. FCM works by applying partition process into certain data to be formed as c-cluster through minimalizing its FCM objective function based on the given limitation[8].

A few research related with FCM had been analyzed by Zhang et al[9] concerning in image clustering and Qiao & Yang[10] related to FCM application in solving the optimal scale function problem by combining dolphin swarm algorithm. Another research in medical sector also done by Yepuganti, Saladi, and Narasimhulu[11] to do a segmentation of tumor disease through FCM application. The data applied in that research are the magnetic resonance (MR) of brain image data which transformed by applying wavelet discreet transformation and feature reduction. PCM and FCM also applied by Upadhyay and Nagpal[12] to reduce the data dimension for EEG data and to classify EGG data into 3 classification, which are awake, SWS, and REM. Other than that, Zhao et al.[13] also did a research related to AE signal classification which analyzed by using PCA and FCM.

Another fuzzy clustering method, FSC also already applied on economic sector by Widodo[14] to form a prediction model of certain brand of products and by Chandar[15] to predict the stock market by applying subtractive clustering with neural fuzzy hybrid approach. FSC also once combined with wavelet to improve the low-cost navigation system with high speed[16]. Another research by Benmouiza and Cheknane[17] also used FSC to classify ANFIS network which applied for predicting the radiation of sun in each hour. In addition to that, FSC also being used by Chen and Li[18] for analyzing the service level of online taxi through PCA combination.

Both of fuzzy clustering method, FCM and FSC, needs certain distance to measure the degree of similarity on certain set of data. In this research, the distance parameter that being applied is Minkowski Chebisev combination distance which invented by Rodrigues. This distance is already used by Rodrigues[19] for classifying process with the application of K-Nearest Neighbors (KNN) and proven to provides high accuracy. Furthermore, this distance also used by Surono and Putri[20] which applied on one of the fuzzy clustering method, FCM, with the combination of PCA. Based on these facts regarding research which already worked before, we decide to do a research concerning in the comparison of the clustering result from FCM and FSC with the application of Minkkowski Chebisev combination distance.

## MATERIAL AND METHODS

### Principal Component Analysis (PCA)

Suppose there are random vector as many as $m$, which are $X_1, X_2, \ldots, X_m$. The main component requirement from variables $X_1, X_2, \ldots, X_m$ are vector of $PC_1, PC_2, \ldots, PC_l$ which acquired through $l < m$, $Var(PC_1) > Var(PC_2) > \cdots > Var(PC_l)$ and $Cov(PC_i, PC_j) = 0$, where. The main component only depends on variants matrix of S from $X_1, X_2, \ldots, X_m$. Generally, the primary component is obtained from the form of linear combination, where the definition of linear combination is:

A vector $w$ is defined as linear combination from vector $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$, if there are scalar $c_1, c_2, \ldots, c_r$, such that

$$\mathbf{w} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_r\mathbf{v_r} \tag{1}$$

Hence, the form of the main component is described as follows:

$$PC_1 = \sum_{j=1}^{m} a_{j1}X_j = a_{11}X_1 + \cdots + a_{m1}X_m$$

$$PC_2 = \sum_{j=1}^{m} a_{j1}X_j = a_{12}X_1 + \cdots + a_{m1}X_m$$

$$\vdots \tag{2}$$

$$PC_m = \sum_{j=1}^{p} a_{jm}X_j = a_{1m}X_1 + \cdots + a_{mm}X_m$$

with

$$a = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{mm} \end{bmatrix}$$

and

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} = [X_1 \quad X_2 \quad \cdots \quad X_m]$$

There are two methods to determine the number of the main components, which are:

First, through proportion of total variant (PTV)

$$PTV = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \times 100\%, \quad k = 1, 2, \ldots, p \quad (3)$$

If the majority of the total variant proportion in m is 80% to 90%, as many as m variables can be connected by the first, second, or third components which have the ability to transform m variables without diminishing its main information[21].

Second, through proportion of cumulative variant (PKV)

$$PKV = \frac{\sum_{j=1}^{k} \lambda_j}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \times 100\% \quad (4)$$

If the majority of the total variant proportion in m is 70% to 80%, as many as m variables can be connected by the first, second, or third components which have the ability to transform m variables without diminishing its main information[22].

## Minkowski Chebysev Combination Distance

The distance combination applied by Rodrigues[19] is Minkowski Chebisev combination distance. The equation of this distance is defined as follows:

$$d_{(w_1, w_2, p)}(x, y) = \left( w_1 \sqrt[p]{\sum_{m=1}^{k} |x_m - y_m|^p} + w_2 max_{m=1}^{k} |x_m - y_m| \right) \quad (5)$$

Where $x_m$ and $y_m$ are the components in $n$ dimension. If the value of $w_1 > w_2$, then the distance will similar to Minkowski and vice versa.

## Fuzzy Subtractive Clustering (FSC)

Based on Kusumadewi and Purnomo[7] ,Fuzzy Subtractive Clustering (FSC) is classified as an unsupervised algorithm where the number of clusters which are going to be formed are still unknown. The basic concept of this method is to determine every data coordinates which have high density towards its neighboring coordinates. The coordinate with the highest number of neighboring data coordinates will be determined as the cluster centroid. Then, the density of this new formed cluster centroid will be reduced and algorithm will choose another coordinate which have the highest neighboring data to be formed as the other new cluster centroid. This process will be initiated until all of the data coordinates are tested.

Two comparison factors will be applied on FSC, which are Accept ratio and Reject ratio. This factors are some certain real fraction numbers ranged from 0 to 1. Accept ratio is the lower limit where certain data coordinates which determined as the cluster centroid candidates are eligible to be defined further as the cluster centroid. While reject ratio is the upper limit where certain data coordinates which determined as the cluster centroid candidates are ineligible to be defined further as the cluster centroid. There are 3 condition probability which can occur during iteration process. The first condition is a condition where ratio > accept ratio, this means that this data coordinate is accepted as the new cluster centroid. The second condition, where reject ratio < ratio ≤ accept, means that this data coordinate will be accepted as the new cluster centroid only if this coordinate is located in certain distance towards another existed cluster centroid (the addition of Ratio and the nearest distance of this data coordinate with another cluster centroid is $(ratio + Mds) \geq 1$). If the addition between the ratio and the farthest distance of this data coordinate with existing cluster centroid is $(ratio + Mds) < 1$, then this data coordinate will not be selected and considered as the new cluster centroid (the potential will be set to 0). Furthermore, the third condition where ratio ≤ reject ratio means that there are no other data coordinates which will be considered to be selected as the new cluster centroid and the iteration process is terminated.

## Fuzzy C-Means (FCM)

Fuzzy clustering is a classification technique where every set of data have the probability to become a member of certain cluster with different degree for each class. There are two type of methods for this fuzzy clustering, the first one is FCM. This technique first introduced by Bezdek in 1981 where certain data set is classified by creating c-number of clusters which already determined before. FCM works by determining the cluster location which will define the average value for each cluster, with inaccuracy on initial condition. Each data coordinate have their own minimum membership degree for each existing cluster. In this condition, the cluster centroid and every level data are iterated over and over, so that the cluster centroid will move to the most suitable position. This iteration process is based on the minimalizing objective function process which positioned into the cluster centroid with the value of each data.

## Validity Index

One of the method to measure the cluster validity is Davies Bouldin Index (DBI). This method works by maximizing the distance between clusters and minimizing the distance between data coordinate in the same cluster. If the distance between clusters is far, then the characteristic similarity between different clusters will be low and the data which belong in certain cluster will tend to have high degree of similarity. Based on DBI, the optimum cluster will have the minimum DBI value (close to 0). The equation for DBI application is described as follows[23]:

$$SSW_i = \frac{1}{m}\sum_{j=1}^{m_i} d(x_j, c_i)$$

$$SSB_{i,J} = d(c_i, c_j)$$

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,J}}$$

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i \neq j}(R_{i,j}) \qquad (6)$$

## RESULT AND DISCUSSION

The measurement process in this study begins with PCA which produces the main component for measuring fuzzy clustering. The data used is hypertension data obtained from a public health center in Yogyakarta. The amount of data as many as 100 data with 5 different variables was applied to this study, meaning that the data used was $100 \times 5$ in size. The 5 variables applied in this research are age $(X_1)$, gender $(X_2)$, systolic pressure $(X_3)$, diastolic pressure $(X_4)$, and body weight $(X_5)$. The data processing in this research is done by applying Software Jupyter Notebook with programming language Phyton.

The PCA calculation is initiated by data normalization to position the data in certain range. Then, matrix covariance is measured to determine whether the relation between variables are exist. The formed matrix covariance then will be analyzed by applying PCA to observe the eigen value for each variable. The formed main component is determined based on the total variance proportion and cumulative proportion. The result of eigen value and variance calculation are illustrated in Table 1.

Table 1. Eigen value and total variance

| Component | Eigen value ($\lambda$) | Variance (%) | Cumulative (%) |
|---|---|---|---|
| 1 | 0.0689 | 30.45 | 30.45 |
| 2 | 0.0569 | 25.11 | 55.56 |
| 3 | 0.0459 | 20.30 | 75.86 |
| 4 | 0.0374 | 16.51 | 92.37 |
| 5 | 0.0173 | 7.62 | 100 |

Table 1 shows that the PKV value in the third component is 75.86%. This value indicates that the existing data set can be presented with 3 components, because the value has exceeded 70%. The eigenvalues indicate the principal component's contribution to the variance of all the original variables analyzed. The graph of the main components and eigenvalues is illustrated in Figure 1. Analysis and interpretation of this graph is necessary before discussing the main results of this study.
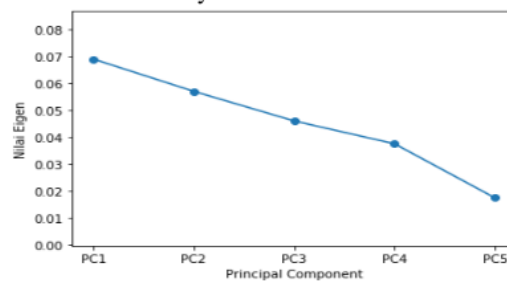
Figure 1. eigen value

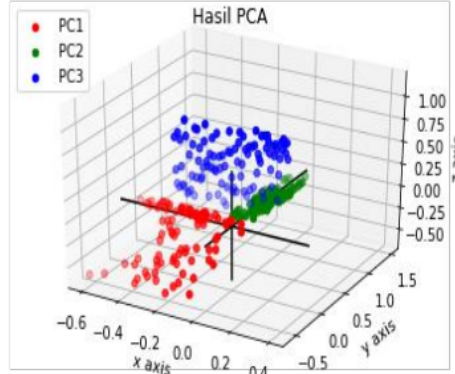3 main components which are formed can be seen from the figure below.



Figure 2. The main component result

Figure 2 visualize the output resulting from the 3 main components, where the red dots, green dots, and blue dots defined as the first, second, and third component respectively. In this research, the FSC process applying squash facto (q) valued 1.25, where this value shows that the radius of the data coordinates around the cluster centroid which will be measured its data potential reduction are valued 1.25. The accept ratio valued 0.8 shows the lower limit value of the data coordinates which selected as the cluster centroid candidate and the reject ratio valued 0.2 shows the upper limit of the data coordinates which selected as the cluster centroid are ineligible to transformed into cluster centroid. Then, the radius which are used in this research are $r = 1.22$, $w_1 = 3$, $w_2 = 2$, $p = 4$. The first step applied in this calculation is transforming the data into fuzzy number by applying exponential membership function Rezaei and Rezaei[24] and resulting the output which is illustrated in table 2.

Table 2. Fuzzy number

| $X_1$ | $X_2$ | $X_3$ |
|---|---|---|
| 1 | 0.1735 | 0.1296 |
| 0.2660 | 0.5406 | 0.4662 |
| ⋮ | ⋮ | ⋮ |
| 0.1584 | 0.4276 | 0.3081 |

Then, the data in Table 2 is processed using FSC method so that the cluster centroid is obtained as below:

$$C_{1.22} = \begin{bmatrix} 1 & 0.1735 & 0.1296 \\ 0.0709 & 0.2443 & 0.6673 \\ 0.3227 & 1 & 0.2524 \end{bmatrix}$$

$C_{1.22}$ shows the cluster centroid formed from the radius 1.22. From the matrix above, the number of rows shows the number of cluster, and the number of column shows the number of applied variables. The first row, second row, and third row illustrate the first, second, and third cluster centroid respectively. Furthermore, the membership value for each data will be calculated by applying this following equation [7].

$$\mu_{k_i} = e^{-\sum_{j=1}^{m}\left(\frac{x_{ij}-C_{kj}}{\sqrt{2}\sigma_j}\right)} \qquad (7)$$

$\mu_{k_i}$ = The membership value of cluster $k$ in data $i$.

$x_{ij}$ = Data $i$ in variable $j$.

The output of the membership value is illustrated in the table 3 below:

Table 3. Membership value

| Value of $\mu$ in cluster 1, 2 and 3 | | |
|---|---|---|
| 1 | 2 | 3 |
| 0.0086 | 0.4878 | 0.2821 |
| 0.5109 | 0.3597 | 0.2297 |
| ⋮ | ⋮ | ⋮ |
| 0.4093 | 0.2445 | 0.2004 |

In table 3, the first data tends to be induced in cluster 2 because the highest membership degree on the first data are located on cluster 2. The second data tends to be induced in cluster 1 because of the highest membership degree on the first data are located on cluster 1, and so on until the 100th data.

Afterwards, the FCM process applied by minimalize this following objective function.

$$min J_m(U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n}(\mu_{ik})^m \left( w_1 \sqrt[p]{\sum_{j=1}^{l}|x_{kj}-v_{ij}|^p} \right.$$
$$\left. + w_2 max_{j=1}^{l}|x_{kj}-v_{ij}| \right) \qquad (8)$$

with limitation
$$\sum_{i=1}^{c}(\mu_{ik})^m = 1 \qquad (9)$$
$$0 < \mu_{ik} < 1$$

with

$J_m(U,V)$ : Objective function U and V

| | |
|---|---|
| $c$ | : Number of cluster |
| $n$ | : Number of data |
| $m$ | : Considered rank, $m \in [1, \infty)$ |
| U | : Matrix partition |
| V | : Cluster centroid matrix |
| $\mu_{ik}$ | : Membership degree of $k$ in cluster $i$ |
| $w_1$ | : Minkowski weight |
| $w_2$ | : Chebysev weight |
| $x_{kj}$ | : Data $k$ in variable $j$ |
| $v_{ij}$ | : Cluster centroid $i$ in variable $j$ |
| $p$ | : Minkowski parameter |

The initial value applied for this calculation are $c = 3$, $m = 2$, iteration maximum = 100, $\varepsilon = 10^{-5}$. For the $w_1, w_2$ and $p$, the value are equal with the value applied for FSC process which are $3, 2$ dan $4$. The partition matrix applied for the calculation is described as follows:

$$U^{(0)} = \begin{bmatrix} 0.1789 & 0.0786 & 0.7425 \\ 0.4496 & 0.3524 & 0.1980 \\ 0.1252 & 0.2773 & 0.5975 \\ \vdots & \vdots & \vdots \\ 0.2688 & 0.0819 & 0.6493 \\ 0.5351 & 0.4130 & 0.0520 \\ 0.7809 & 0.1588 & 0.0602 \end{bmatrix}$$

Then, the first iteration process resulting in the cluster centroid described as follows:

$$v_{ij}^{(1)} = \begin{bmatrix} -0.0550 & 0.0775 & 0.0152 \\ 0.0798 & -0.1368 & 0.0179 \\ 0.1184 & -0.0334 & -0.0164 \end{bmatrix}$$

Where $i = 1,2,3$ dan $j = 1,2,3$.

The objective function acquired from the first iteration is

$$F^{(1)} = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^m \left( w_1 \sqrt[p]{\sum_{j=1}^{l} |x_{kj} - v_{ij}|^p} \right. $$
$$\left. + w_2 max_{j=1}^{l} |x_{kj} - v_{ij}| \right)$$
$$= 330,3307.$$

After that, the iteration process are continuously iterated until $t > maxiter$ or $\left| F_{FCMMC}^{(t)}(U,V) - F_{FCMMC}^{(t-1)}(U,V) \right| < \varepsilon$. If these condition is meet, then the iteration process will be terminated. The cluster centroid acquired from this process are described in the matrix below.

$$v_{ij}^{(100)}$$
$$= \begin{bmatrix} -0.0608 & 0.0816 & 0.0083 \\ 0.0222 & -0.0826 & 0.0078 \\ 0.0406 & -0.0022 & -0.0161 \end{bmatrix}, \begin{matrix} i & = & 1,2,3 \\ j & = & 1,2,3. \end{matrix}$$

As for the new partition matrix which acquired when $t = 100$, the output is described in this following matrix.

$$U^{(0)} = \begin{bmatrix} 0.3496 & 0.3142 & 0.3364 \\ 0.3386 & 0.3340 & 0.3273 \\ 0.2892 & 0.3760 & 0.3348 \\ \vdots & \vdots & \vdots \\ 0.3341 & 0.3208 & 0.3451 \\ 0.3715 & 0.2997 & 0.3289 \\ 0.3319 & 0.3288 & 0.3394 \end{bmatrix}$$

Furthermore, the cluster centroid resulted from FSC and FCM process will be evaluated by applying equation (6). The SSW and SSB value resulted from FSC are:

$$\begin{aligned} SSW_1 &= 3.5847 \\ SSW_2 &= 5.6233 \\ SSW_3 &= 6.2458 \\ SSB_{12} &= 2.9925 \\ SSB_{13} &= 3.0374 \\ SSB_{23} &= 2.7893 \end{aligned}$$

While the SSW and SSB value resulted from FCM are:

$$\begin{aligned} SSB_{12} &= 6.3353 \\ SSB_{13} &= 7.5549 \\ SSB_{23} &= 8.5845. \\ SSW_1 &= 5.4171 \\ SSW_2 &= 3.9744 \\ SSW_3 &= 7.6456. \end{aligned}$$

Based on equation (6) the values for the FCM and FSC methods are obtained which are shown in Table 3.

**Table 3.** DBI Value

| | PCA | NonPCA |
|---|---|---|
| FCM | 1.6468 | 2.0406 |
| FSC | 3.9164 | 0.5998 |

In Table 1, the FCM method produces a DBI value of 1.6468 using PCA and 2.0406 without using PCA. Meanwhile, the FSC method produces a DBI value of 3.9164 using PCA and 0.5998 without using PCA.

## CONCLUSION

This research combines PCA and fuzzy clustering methods. PCA is applied for dimensional reduction. Through this research, PCA produced 3 new variables which were the new main components of the original 5

variables. Then the results from the PCA will be used as input parameters for the FSC and FCM methods. The distance used for fuzzy clustering is the combination distance of Minkowski Chebisev. Furthermore, the cluster evaluation process is carried out by applying DBI, where the DBI value for FSC using PCA is 3.9164 and without PCA is 0.5998. for FCM it is 1.6468.

Meanwhile, the DBI value for FCM using PCA is 1.6468 and without PCA is 2.0406. Therefore, in this case, if you use the FCM method, the existing data set is reduced first using PCA, but if the FSC method is used, the DBI value will be even better if you don't use PCA first.

## BIBLIOGRAPHY

[1]     Y. Aït-Sahalia and D. Xiu, "Principal Component Analysis of High-Frequency Data," *J. Am. Stat. Assoc*., vol. 114, no. 525, pp. 287–303, Jan. 2019, doi: 10.1080/01621459.2017.1401542.

[2]     A. Matuschek *et al*., "Analysis of parathyroid graft rejection suggests alloantigen-specific production of  nitric oxide by iNOS-positive intragraft macrophages.," *Transpl. Immunol*., vol. 21, no. 4, pp. 183–191, Sep. 2009, doi: 10.1016/j.trim.2009.04.004.

[3]     I.T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer-Verlag, 2002.

[4]     G. Gan, C. Ma, and J. Wu, *Data Clustering:     Theory,    Algorithms,    and Applications*. Society for Industrial and Applied Mathematics, 2007.

[5]     A. C. Rencher and W. F. Christensen, *Methods of Multivariate Analysis*. Wiley, 2012.

[6]     J. S. R. Jang, C. T. Sun, and E. Mizutani, "Neuro-Fuzzy    and    Soft    Computing-A Computational  Approach  to  Learning  and Machine Intelligence [Book Review]," *IEEE Trans. Automat. Contr*., vol. 42, no. 10, pp. 1482–1484,    Oct.    1997,    doi: 10.1109/TAC.1997.633847.

[7]     S. Kusumadewi  and  H. Purnomo, "Aplikasi  Logika  Fuzzy  untuk  Pendukung Keputusan, vol. II," *Yogyakarta Graha Ilmu*, 2013.

[8]     J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Boston, MA: Springer US, 1981.

[9]     L. Zhang, M. Luo, J. Liu, Z. Li, and Q. Zheng,  "Diverse  fuzzy  c-means  for  image clustering," *Pattern Recognit. Lett*., vol. 130, pp. 275–283,      2020,      doi: https://doi.org/10.1016/j.patrec.2018.07.004.

[10]     W. Qiao and Z. Yang, "An Improved Dolphin  Swarm  Algorithm  Based  on  Kernel Fuzzy  C-Means  in  the  Application  of  Solving the Optimal Problems of Large-Scale Function," *IEEE Access*, vol. 8, pp. 2073–2089, 2020, doi: 10.1109/ACCESS.2019.2958456.

[11]     Y. Karuna, S. Saladi, and C. V. Narasimhulu, "Segmentation of tumor using PCA based modified fuzzy C means algorithms on MR brain images," *Int. J. Imaging Syst. Technol*.,     vol.     30,     Jun.     2020,     doi: 10.1002/ima.22451.

[12]     P. Upadhyay and C. Nagpal, "PCA-Aided FCM Identifies Stressful Events of Sleep EEG Under Hot Environment," *IETE J. Res*., pp. 1–14,       Jun.       2020,       doi: 10.1080/03772063.2020.1782273.

[13]     G. Zhao, L. Zhang, C. Tang, W. Hao, and Y. Luo, "Clustering of AE signals collected during  torsional  tests  of  3D  braiding  composite shafts using PCA and FCM," *Compos. Part B Eng*.,  vol.  161,  pp.  547–554,  2019,  doi: https://doi.org/10.1016/j.compositesb.2018.12.145.

[14]     I.  D.  Widodo,  "Fuzzy  subtractive clustering  based  prediction  model  for  brand association analysis," *MATEC Web Conf*., vol. 154,   p.   01082,   Feb.   2018,   doi: 10.1051/matecconf/201815401082.

[15]     S. K. Chandar, "Stock market prediction using subtractive clustering for a neuro fuzzy hybrid approach," *Cluster Comput*., vol. 22, no. S6,   pp.   13159–13166,   Nov.   2019,   doi: 10.1007/s10586-017-1321-6.

[16]     E. S. Abdolkarimi and M. R. Mosavi, "Wavelet-adaptive  neural  subtractive  clustering fuzzy inference system to enhance low-cost and high-speed INS/GPS navigation system," *GPS Solut*., vol. 24, no. 2, p. 36, Apr. 2020, doi: 10.1007/s10291-020-0951-y.

[17]     K.  Benmouiza  and  A.  Cheknane, "Clustered ANFIS network using fuzzy c-means,

subtractive clustering, and grid partitioning for hourly solar radiation forecasting," *Theor. Appl. Climatol.*, vol. 137, no. 1–2, pp. 31–43, Jul. 2019, doi: 10.1007/s00704-018-2576-4.

[18]     H. Chen and C. Li, "Comprehensive Service Level Analysis of Online Taxi Drivers Based on Fuzzy Clustering Combined with Principal Component Analysis," in *Proceedings of the 2019 4th International Conference on Humanities Science and Society Development (ICHSSD 2019)*, 2019, pp. 577–583, doi: 10.2991/ichssd-19.2019.116.

[19]     E. O. Rodrigues, "Combining Minkowski and Chebyshev: New distance proposal and survey of distance metrics using k-nearest neighbours classifier," *Pattern Recognit. Lett.*, vol. 110, pp. 66–71, 2018, doi: https://doi.org/10.1016/j.patrec.2018.03.021.

[20]     S. Surono and R. D. A. Putri, "Optimization of Fuzzy C-Means Clustering Algorithm with Combination of Minkowski and Chebyshev Distance Using Principal Component Analysis," *Int. J. Fuzzy Syst.*, 2020, doi: 10.1007/s40815-020-00997-5.

[21]     R. A. Johnson and D. W. Wichern, "Applied multivariate statistical analysis," *Statistics (Ber).*, vol. 621, no. 10, p. 10, 2015.

[22]     V. Gaspersz, "Teknik analisis dalam penelitian percobaan," *Tarsito. Bandung*, vol. 718, 1992.

[23]     R. Gustriansyah, N. Suhandi, and F. Antony, "Clustering optimization in RFM analysis Based on k-Means," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 18, no. 1, p. 470, Apr. 2020, doi: 10.11591/ijeecs.v18.i1.pp470-477.

[24]     K. Rezaei and H. Rezaei, "New distance and similarity measures for hesitant fuzzy soft sets," *Iran. J. Fuzzy Syst.*, vol. 16, no. 6, pp. 159–176, 2019.

# 254-99Z_Article Text-1053-1-6-20210301.docx

7   P Laras Sakti Cahyaningrum, Nursyiva Irsalinda. "Indonesian community welfare levels clustering using the fuzzy subtractive clustering (FCM) method", Journal of Physics: Conference Series, 2019
Crossref

39 words — 1%

8   Sana Iqbal, Mohammad Ayyub. "Improved Performance of Fuzzy Logic Controller to Control Dynamical Systems: A Comparative Study", 2018 International Conference on Computational and Characterization Techniques in Engineering & Sciences (CCTES), 2018
Crossref

38 words — 1%

9   Anandkumar Balasubramaniam, Malik Junaid Jami Gul, Varun G. Menon, Anand Paul. "Blockchain For Intelligent Transport System", IETE Technical Review, 2020
Crossref

30 words — 1%

10  Babahadj Abdeljabar, Lakhdar Rahmani, Bellaoui Mebrouk, Bouchouicha Kada. "Estimating daily diffuse solar radiation using neuro-fuzzy, fuzzy C-means and subtractive clustering", Proceedings of the 10th International Conference on Information Systems and Technologies, 2020
Crossref

28 words — 1%

11  Hongzhi Jia, Sheng Liang, Shuqin Lou, Xinzhi Sheng. "A $k$ -Nearest Neighbor Algorithm-Based Near Category Support Vector Machine Method for Event Identification of $\varphi$ -OTDR", IEEE Sensors Journal, 2019
Crossref

27 words — 1%

12  www.warse.org
Internet

27 words — 1%

13  uzjournals.edu.uz
Internet

24 words — 1%

**14** Cao, Guoyan, John Michelini, Karolos Grigoriadis, Behrouz Ebrahimi, and Matthew A. Franchek. "Cluster-based correlation of severe braking events with time and location", 2015 10th System of Systems Engineering Conference (SoSE), 2015.

23 words — 1%

Crossref

EXCLUDE QUOTES         OFF
EXCLUDE BIBLIOGRAPHY   OFF

EXCLUDE MATCHES        < 1%