Available online at :
http://ejournal.amikompurwokerto.ac.id/index.php/telematika/

Telematika

Accreditated SINTA "2" Kemenristek/BRIN, No. 85/M/KPT/2020

Comparison of Industrial Business Grouping using Fuzzy C-Means and Fuzzy Possibilistic C-Means Methods

Mega Lestari[1], Dwi Kartini[2*], Irwan Budiman[3], Mohammad Reza Faisal[4] and Muliadi[5]

[1,2,3,4,5] Computer Science, Faculty of Mathematics and Natural Sciences
Lambung Mangkurat University, Banjarmasin, Indonesia
E-mail: megalstr2@gmail.com[1], dwikartini@ulm.ac.id[2], irwan.budiman@ulm.ac.id[3]
reza.faisal@ulm.ac.id[4], muliadi@ulm.ac.id[5]

ARTICLE INFO

History of the article:
Received July 8, 2023
Revised August 30, 2023
Accepted September 13, 2023

Keywords:
Industrial Business,
Clustering,
Fuzzy C-Means,
Fuzzy Possibilistic C-Means,
Cluster Validity Index

Correspondece:
E-mail: dwikartini@ulm.ac.id

ABSTRACT

The industrial business sector plays a role in the development of the economic sector in developing countries such as Indonesia. In this case, many industrial businesses are growing, but the data has not been processed or analyzed to produce important information that can be processed into knowledge using data mining. One of the data mining techniques used in this research is data grouping, or clustering. This research was conducted to determine the comparison results of the Cluster Validity Index on Fuzzy C-Means and Fuzzy Possibilistic C-Means methods for clustering industrial businesses in Tanah Bumbu Regency. In each process, 5 trials were conducted with the number of clusters, namely 3, 4, 5, 6, and 7, and for the attributes used: Male Labor, Female Labor, Investment Value, Production Value, and BW/BP Value. Furthermore, this study will evaluate the Cluster Validity Index, namely the Partition Entropy Index, Partition Coefficient index, and Modified Partition Coefficient Index. This research provides the best performance results in the Fuzzy C-Means method with the results of the Cluster Validity Index on the Partition Entropy Index of 0.21566, Partition Coefficient Index of 0.88078, and Modified Partition Coefficient Index of 0.82117, and the best number of clusters is 3 with the labels of low competitive industry clusters, medium competitive industry clusters, and highly competitive industry clusters.

INTRODUCTION

The industrial business sector plays a role in the development of the economic sector in developing countries such as Indonesia. Industrial businesses also make a large contribution to regional income and play a role in absorbing labor in the region. In this case, many industrial businesses are growing, but the data has not been processed or analyzed to produce important information that can be processed into knowledge using data mining. One of the data mining techniques used in this research is data grouping, or clustering.

The clustering method is the process of dividing data into groups based on their level of similarity. Data that has similar characteristics will gather in one group (Putri et al., 2022). Each clustering method has its own advantages and disadvantages. Optimal cluster results can be influenced by the clustering method used, the characteristics of the dataset, and the number of clusters used. Therefore, it is necessary to be precise when using clustering methods in order to get optimal cluster results.

This research will focus on the partition clustering method, namely Fuzzy. One of the algorithms, Fuzzy C-Means, is an algorithm that combines fuzzy principles with the K-Means method. Fuzzy C-Means is a data clustering technique where the presence of each data point in a group is determined by the degree of membership of the data. Research by (Firdaus et al., 2021) determined the best method for mapping crime-prone areas in Semarang City and found that the Fuzzy C-Means method is better than the K-Means method with a percentage of 71.23%. Nidyashofa & Istiawan (2017) mentioned that Fuzzy C-Means has advantages in placing cluster centers that are more precise than other cluster methods. However, Fuzzy C-Means has a weakness in overcoming uncertainty and providing a clear interpretation of the resulting partition.

There are several previous studies related to the Fuzzy C-Means and Fuzzy Possibilistic C-Means methods, including those conducted by Thilagaraj & Sengottaiyan (2019). With data on students who want to work in the software industry, a comparison of methods is carried out to determine the best method based on performance. The best method based on performance using R-Tool is the Fuzzy Possibilistic C-Means method with the best cluster is 4. As well as research by Apsari et al. (2020) using student performance data and implementing two methods, namely Fuzzy C-Means and Possibilistic C-Means. Resulting in the Possibilistic C-Means method as a better method with the best cluster is 4.

Fuzzy C-Means has long been used to cluster data with fuzzy group members, reflecting the degree of uncertainty in the clustering. On the other hand, Fuzzy Possibilistic C-Means is used as an alternative with its ability to handle uncertainty more expansively through the concept of possibilistic membership functions. In addition to the clustering method used, there is a method to evaluate the cluster performance results obtained from Fuzzy C-Means and Fuzzy Possibilistic C-Means. Knowing this, it is necessary to have a Cluster Validity index, which is a validity measure, to find the optimal number of clusters that can fully explain the data structure.

Research by Widiyanto (2019) measured the quality of clustering with the Cluster Validity Index, namely the Partition Coefficient Index, Classification Entropy Index, Partition Entropy Index, Fukuyama Sugeno Index, Xie Beni Index, Modified Partition Coefficient Index, and Exponential Separation Index. It is obtained that the use of the Cluster Validity Index is proven to be able to measure the Fuzzy C-Means method. Research by Handoyo et al. (2019) conducted a comparison of Fuzzy C-Means and Subtractive Fuzzy C-Means methods on public health facility data, resulting in the best method being Fuzzy C-Means. And the best cluster is 3, based on the Cluster Validity Index with a Partition Coefficient Index value of 0.74749, a Modified Partition Coefficient Index value of 0.62123, and a Classification Entropy value of 0.15055. Furthermore, Herlinda et al. (2021) used first-level health facility data with the Fuzzy C-means method. Provide conclusions on the Cluster Validity Index value, namely the Partition Entropy Index, namely 0.50002 and the Partition Coefficient Index, namely 0.99998 with the best cluster is 2.

In research by Rouza & Fimawahib (2020) based on small and medium business data for data clustering using Fuzzy C-Means. The best cluster is obtained, namely 3 clusters with UAT (User Acceptance Test), namely the average validation value is almost close to 1, this shows that Fuzzy C-Means Clustering has a high accuracy rate of 80-90%. Research by Firdaus et al. (2021) using crime data to classify data using Fuzzy C-Means by looking at the results based on the Cluster Validity Index, namely the Partition Coefficient Index value is 0.81800 with the best cluster is 5. As well as researchers by Amaliyah et al. (2022) using data on the distribution of contraceptive users, it is known that Fuzzy C-Means is the best method seen from the highest Modified Partition Coefficient value in each cluster which is 0.46407 with the best cluster is 2.

Based on the explanation described above, this study wanted to know the comparison of the Fuzzy C-Means and Fuzzy Possibilistic C-Means methods in industrial business data clustering, using the cluster validity index evaluation. The cluster validity index used is based on previous research, namely Partition Entropy Index, Partition Coefficient Index, and Modified Partition Coefficient Index. The main objective of this comparison is to determine the clustering method with the best performance and the most suitable clusters for grouping industrial business data.

RESEARCH METHODS

There are research methods that will be carried out to ensure the smooth running of the research in line with the objectives, as well as to determine the method that has the best performance for grouping industrial business data. The flow of research methods can be seen in Figure 1.

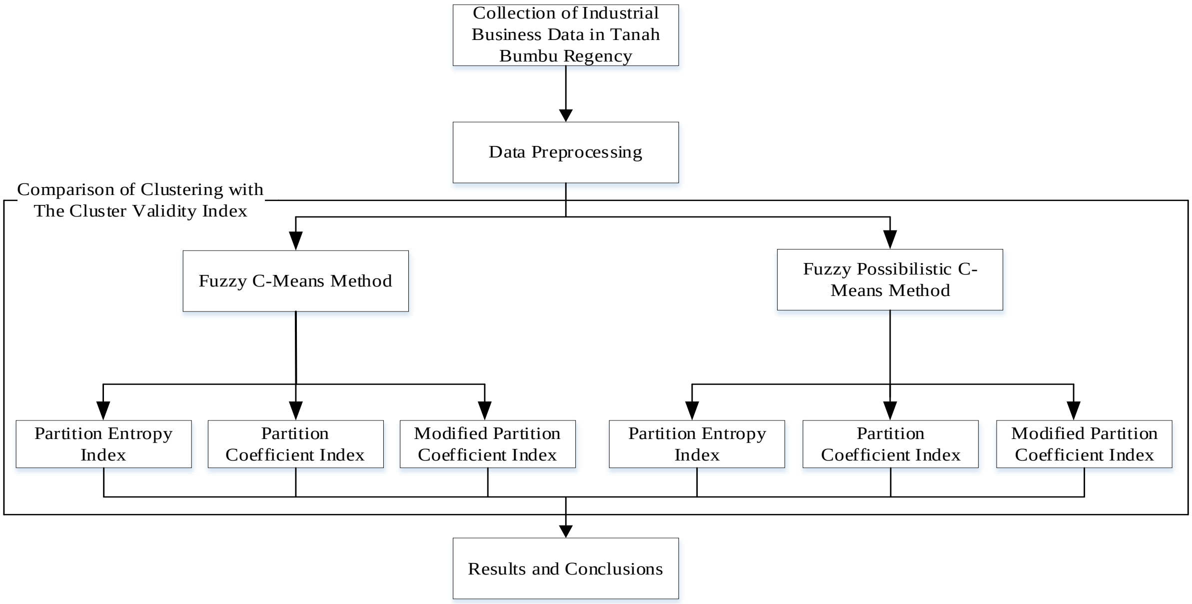

Figure 1. Research Flow

1. Data Collection

This research uses industrial business data in Tanah Bumbu Regency in 2022 taken from the South Kalimantan Provincial Industry Office, consisting of 31 attributes and totaling 2,947 data points. The data will be grouped into industrial businesses. The following is an example of data that will be used in research, as seen in Table 1.

Table 1. Examples of Industrial Business Data

Number	*Nama of Company*	*Name of Owner*	…	...	*Access to Financing*
1	Cv. Mng	Sy	…	…	Personal
2	Ud.Ks	Ks	…	…	Personal
3	Ud.K U	Gfr	…	…	Personal
…	…	…	…	…	…
…	…	…	…	…	…
…	…	…	…	…	…
2945	Rmw	Padasuka	…	…	Personal
2946	The Ment	Ar	…	…	Personal
2947	Salsa	Hely	…	…	Personal

In the example of industrial business data in Table 1, which consists of 31 attributes, it is used as information for each industrial business data consisting of character and numeric data types. In this study, industrial business data of the numerical data type consists of male workers, female workers, investment values, production values, and BB/BP values.

2. Data Preprocessing

This research consists of several stages. In the first stage, the data collection process was obtained from the South Kalimantan Provincial Office of Industry, and the data was from all industrial business sectors in Tanah Bumbu Regency in Excel format, which can be seen in Table 1. The second stage is data preprocessing, namely data preparation with several steps will do is:

a. Data cleaning is used to select data and remove data that has the potential to reduce accuracy. In this study, fill in data manually for data with a value of "-" with "0" in the attributes of male workers and female workers, and delete data manually if there is missing value data or data that does not match the attributes investment value, production value, and BB/BP value. There are 18 missing values and data that are not suitable for research.
b. Data integration to combine data from various sources into one larger data unit. In this study, data was not combined into a new dataset because this study only used one data source, namely industrial data from the South Kalimantan Provincial Office of Industry in 2022.
c. Data reduction, or data reduction, is to reduce the amount of data taken without affecting the data analysis process. In this study, we reduced the attributes that did not affect the industrial business grouping process, and the attributes to be used were male workers, female workers, investment values, production values, and BB/BP values.
d. Data transformation is the process of changing the data structure, format, or values so as to produce a dataset that is suitable for the data mining process. In this case, the investment value, production value, and BB/BP value attributes are divided by 1000 to get a smaller value. For example, in data 1, the value of the investment value attribute is 200.000 to 200, the value of the production value attribute is 300.000 to 300, the value of the BB/BP is 130.000 to 130, and so on for all the data that will be used to facilitate the analysis process and data grouping.

The results of data preprocessing can be seen in Table 2, which was obtained after data preprocessing, namely 2.929 rows of data stored in CSV format and using 5 attributes: male workers, female workers, investment value, production value, and BB/BP value.

Table 2. Samples Data after Preprocessing Data

Number	*Male Workers*	*Female Workers*	*Investment Values*	*Production Values*	*BB/BP Values*
1	10	10	200	300	130
2	10	10	200	300	150
3	10	10	150	200	100
4	10	10	200	300	120
5	7	8	200	300	127
…	…	…	…	…	…
…	…	…	…	…	…
…	…	…	…	…	…
…	…	…	…	…	…
…	…	…	…	…	…
2925	0	1	5	18	10
2926	0	1	5	14,4	7
2927	0	1	4	8,5	4,5
2928	1	1	5	36	20
2929	0	1	5	7,2	3,5

3. Comparison of Clustering with The Cluster Validity Index

Furthermore, data mining modeling is grouped using the Fuzzy C-Means and Fuzzy Possibilistic C-Means methods, and in Table 3 are the parameters that will be used for research.

Table 3. Parameters for Research

Parameters	*Method used*	
	Fuzzy C-Means	*Fuzzy Possibilistic C-Means*
Number of clusters (*c*)	3,4,5,6,7	3,4,5,6,7
Weighting power (*w*)	2	2
PCM Weighting power (η)	-	2
Maximum iteration (MaxIter)	1000	1000
Smallest error (ξ)	0,001	0,001
Initial objective function (P_o)	0	0
Initial iteration (*t*)	1	1

As for the steps of the fuzzy C-Means method, they are as follows (Thilagaraj & Sengottaiyan, 2019):

1. Enter data in the form of an n x m matrix, which has a function to determine the amount of data and the number of attributes or variables for each piece of data that will be used.

$$X_{ij} = \begin{bmatrix} X_{11} & \cdots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nm} \end{bmatrix} \quad (1)$$

Informations :

n = number of data samples

m = attribute of each data

X_{ij} = sample data to-i (i=1,2,..,n), attribute to-j (j=1,2,..,m)

2. Determine:
   a. Number of cluster = *c* (≥2);
   b. Weighting power = *w* (>1);
   c. Maximum iteration = MaxIter;
   d. Smallest error = ξ;
   e. Initial objective function = P_o =0;
   f. Initial iteration = *t* =1;
3. Generating random numbers (μ_{ik}) with $i = 1,2, \dots n$; $k = 1,2, \dots, n$; and c as elements of the initial partition matrix U, or relative uniqueness matrix. Here is the form of the initial partition matrix:

$$U_{ik} = \begin{bmatrix} \mu_{11}(X_1) & \cdots \mu_{1c}(X_1) \\ \vdots & \ddots \quad \vdots \\ \mu_{11}(X_n) & \cdots \mu_{nc}(X_n) \end{bmatrix} ; 1 \le i \le n; 1 \le k \le c \quad (2)$$

The sum of each column element value in a row is 1.

4. Calculating the cluster center is indicated by V_{kj} and the following equation:

$$V_{kj} = \frac{\sum_{i=1}^{n}((\mu_{ik})^{w} * X_{ij})}{\sum_{i=1}^{n}(\mu_{ik})^{w}} \quad (3)$$

5. Calculating the objective function, *Pt* with the following equation:

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} \left(\left[\sum_{j=1}^{m} (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \quad (4)$$

6. Calculation the change in the partition matrix with the following equation:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^2\right]^{\frac{-1}{w-1}}} \quad (5)$$

7. Check stop condition:
   a. If ($|P_t - P_{t-1}| < \xi$) or ($t > MaxIter$) then stop;
   b. If not, $t=t+1$, repeat step 4.
8. Determine the members in each cluster.
   With data objects, they are said to be included in a cluster if themembership value is close to 1 or the largest.

Meanwhile, the steps for the Fuzzy Possibilistic C-Means method are as follows (Thilagaraj & Sengottaiyan, 2019):

1. Enter data for group X in the form of an n x m matrix, which has the function of determining the amount of data and the number of attributes or variables for each data set to be used.

$$X_{ij} = \begin{bmatrix} X_{11} & \cdots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nm} \end{bmatrix} \quad (6)$$

   Informations:
   n = number of data samples
   m = attribute of each data
   X_{ij} = sample data to-i (i=1,2,..,n), attribute to-j (j=1,2,..,m)
2. Determine:
   a. Number of clusters = c (≥2);
   b. Weighting power = w (>1) dan η (>1);
   c. Maximum iteration = MaxIter;
   d. Smallest error = ξ;
   e. Initial objective function = $P_o = 0$;
   f. Initial iteration = t = 1;
3. Generating relative uniqueness matrix (μ_{ik}) and cluster centers (V_{kj}) on Fuzzy C-Means to calculate absolute uniqueness matrix as follows:

$$t_{ik} = \begin{bmatrix} t_{11} & \cdots & t_{1c} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{nc} \end{bmatrix} \quad (7)$$

   With the matrix elements as follows:

$$t_{ik} = \frac{\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^2\right]^{\frac{-1}{\eta-1}}}{\sum_{i=1}^{n}\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^2\right]^{\frac{-1}{\eta-1}}} \quad (8)$$

4. Fix the cluster center with the following equation:

$$v_{kj} = \frac{\sum_{i=1}^{n}(\mu_{ik}^{w}+t_{ik}^{\eta})x_{ij}}{\sum_{i=1}^{n}(\mu_{ik}^{w}+t_{ik}^{\eta})} \quad (9)$$

5. Calculating the objective function at iteration = t with the following equation:

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^2\right](\mu_{ik}{}^{w}t_{ik}{}^{\eta})\right) \quad (10)$$

6. Fixed the relative uniqueness matrix (μ_{ik}) with the following equation:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^2\right]^{\frac{-1}{w-1}}}{\sum_{i=1}^{c}\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^2\right]^{\frac{-1}{w-1}}} \quad (11)$$

7. Fixed the absolute uniqueness matrix (t_{ik}) with the following equation:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}(x_{ij}-v_{kj})^2\right]^{\frac{-1}{\eta-1}}}{\sum_{i=1}^{c}\left[\sum_{j=1}^{m}(x_{ij}-v_{kj})^2\right]^{\frac{-1}{\eta-1}}} \tag{12}$$

8. Checking stop condition:
   a. If ($|P_t - P_{t-1}| < \xi$) or ($t > MaxIter$) the stop;
   b. If not, $t=t+1$, repeat step 4.
9. Determine the members in each cluster.

   With data objects, they are said to be included in a cluster if the membership value is close to 1 or the largest.

After clustering, the final stage of evaluation is using the cluster validity index, which is a validity measure used to find the optimal number of clusters. The cluster validity index used includes the Partition Entropy Index, Partition Coefficient Index (Mashfuufah & Istiawan, 2018). The Partition Entropy Index (PEI) method is defined by the equation:

$$PEI = -\frac{1}{n}\sum_{k=1}^{c}\sum_{i=1}^{n}\mu_{ik}\ln(\mu_{ik}) \tag{13}$$

The Partition Coefficient Index (PCI) method is defined by the equation:

$$PCI = \frac{1}{n}\sum_{k=1}^{c}\sum_{i=1}^{n}\mu_{ik}{}^2 \tag{14}$$

Where:

n = many research objects

c = many clusters

μ_{ik} = memberships value of the i object in the j cluster

Partition Entropy Index is an evaluation that measures the level of vagueness of the cluster partition; the optimal cluster is obtained if the value obtained is smaller (closer to 0). While the Partition Coefficient Index only evaluates the membership degree value, the optimal cluster is obtained if the value obtained is getting bigger (closer to 1) (David et al., 2020).

The Modified Partition Coefficient Index (MPCI) method is defined by the equation (Amirah et al., 2017):

$$MPC = 1 - \frac{c}{c-1}(1 - PCI) = 1 - \frac{c}{c-1}\left(1 - \left(\frac{1}{n}\sum_{k=1}^{c}\sum_{i=1}^{n}\mu_{ik}{}^2\right)\right) \tag{15}$$

Where:

n = many research objects

c = many clusters

μ_{ij} = memberships value of the i object in the j cluster

The optimal cluster is obtained if the Modified Partition Coefficient Index value obtained is getting bigger (closer to 1). The smaller or closer the MPCI value to 0, the more blurred the accuracy (Amirah et al., 2017).

4. Result and Conclusion

This stage provides conclusions from the results of research that has been done and describes the characteristics of the optimal cluster.

RESULTS AND DISCUSSION

Grouping or clustering is done using two methods, namely Fuzzy C-Means and Fuzzy Possibilistic C-Means. In the experiment, 5 experiments were carried out with the number of clusters 3, 4, 5, 6, and 7, and using the parameters in Table 3. The following in Table 4 includes the overall results for the number of iterations, objective function, and time using fuzzy C-Means.

Table 4. Iteration Results, Objective Functions, and Time using Fuzzy C-Means

Number of clusters	*Total Iterations*	*Objective Functions*	*Time*
3	56	21249960	17,75
4	61	14369840	24,27
5	58	11109928	26,02
6	61	7373888	34,76
7	106	6205906	70,72

As well, Table 5 includes the overall results of the number of iterations, objective function, and time using Fuzzy Possibilistic C-Means.

Table 5. Iteration Results, Objective Functions, and Time Using Fuzzy Possibilistic C-Means

Number of clusters	*Total Iterations*	*Objective Functions*	*Time*
3	8	42058885	2,57
4	9	52145711	3,62
5	9	69359798	4,58
6	10	71395006	6,56
7	10	72110938	7,2

From both Table 4 and Table 5, overall, the data shows variations in performance and results between the various groupings. However, in order to make better conclusions and choose the best performance method and the optimal number of clusters, a better analysis is needed using an evaluation method in accordance with the objective, namely to obtain the best method for grouping industrial business data. So, this study evaluates using the cluster validity index to determine the best performance method and the best number of clusters. In the cluster validity index using the results of the Partition Entropy Index, Partition Coefficient Index, and Modified Partition Coefficient Index, which were carried out on the Fuzzy C-Means and Fuzzy Possibilitic C-Means methods. The results of the Partition Entropy Index values can be seen in Figure 2 for the two clustering methods.

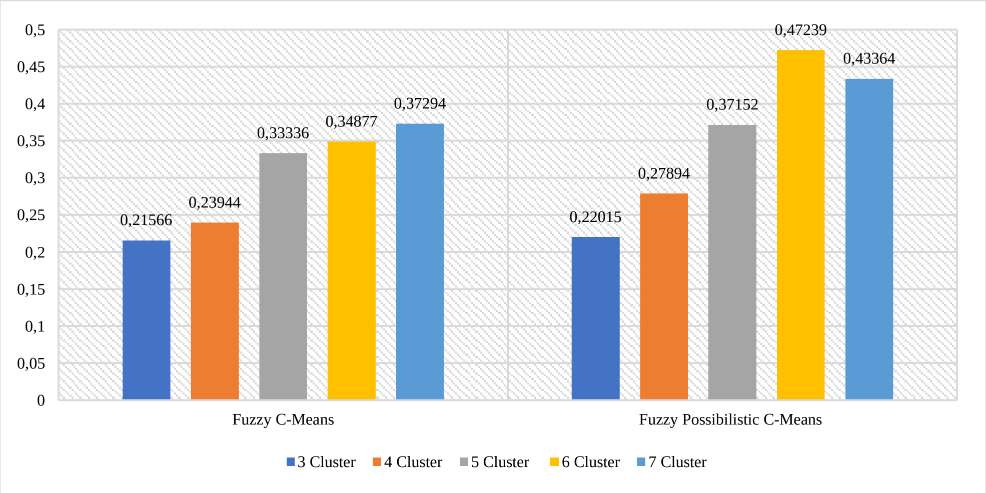

Figure 2. Partition Entropy Index Result

From the results of the Partition Entropy Index values between Fuzzy C-Means and Fuzzy Possibilistic C-Means, it shows that the quality of grouping based on the number of clusters is good if the value obtained is close to small or close to 0, then the results for Fuzzy C-Means are obtained for the number of clusters is 3, with a value of 0.21566. Whereas the Fuzzy Possibilistic C-Means results yield a value of 0.22015, with the best number of clusters being 3. The conclusion is that the Fuzzy C-Means method is better with optimal clusters is 3. The results of the Partition Coefficient Index values can be seen in Figure 3 for the two clustering methods.

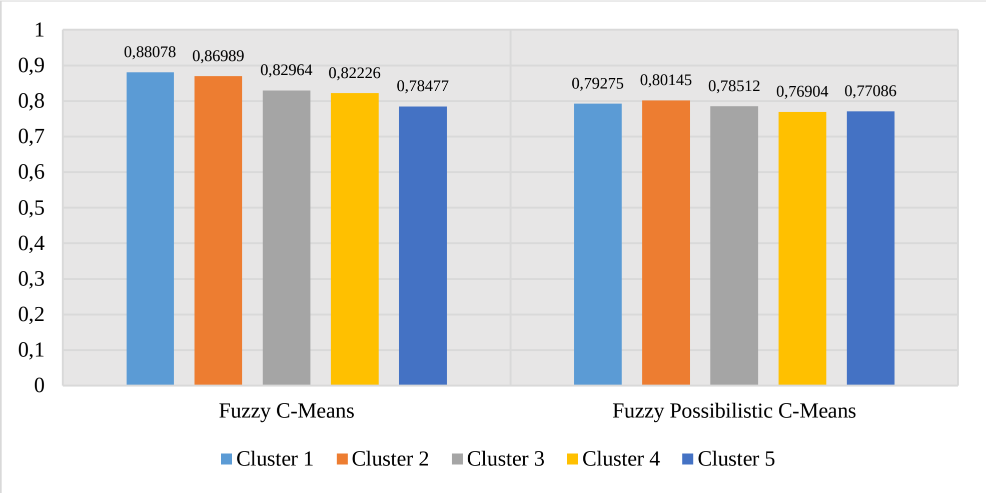

Figure 3. Partition Coefficient Index Results

From the results of the Partition Coefficient Index values between Fuzzy C-Means and Fuzzy Possibilistic C-Means, it shows that the quality of grouping based on the number of clusters is good. If the value obtained is close to large or close to 1, then the results for Fuzzy C-Means are obtained for the best number of clusters. is 3, with a value of 0.88078. Whereas the Fuzzy Possibilistic C-Means results yield a value of 0.80145, with the best number of clusters being 4. The conclusion is that the Fuzzy C-Means method is better with optimal clusters is 3. The results of the Partition Coefficient Index values can be seen in Figure 4 for the two clustering methods.

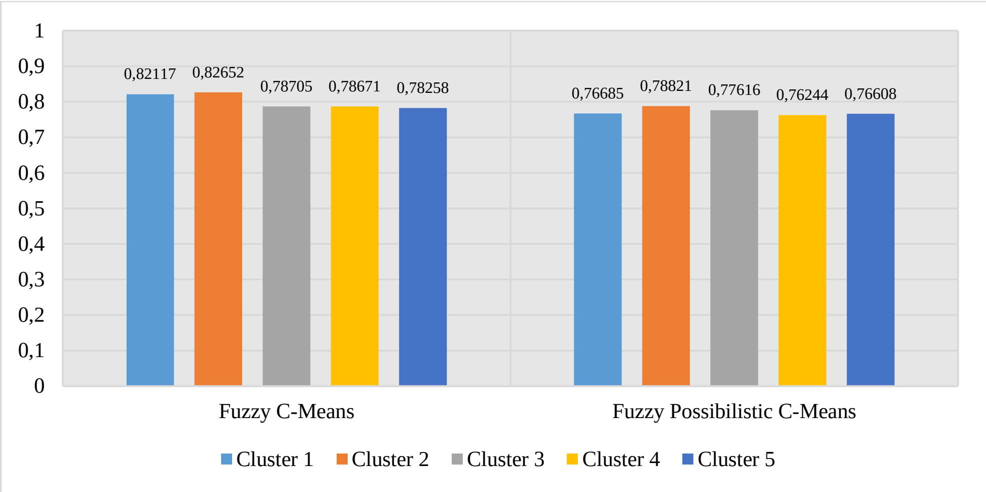

Figure 4. Modified Partition Coefficient Index Results

Finally, the results of the Modified Partition Coefficient Index values between Fuzzy C-Means and Fuzzy Possibilistic C-Means show variations in the quality of grouping based on the number of clusters that are stated to be good. If the values obtained are close to large or close to 1, then the results for Fuzzy C-Means are obtained for the sum, and the best cluster is 4 with a value of 0.82652. Whereas the Fuzzy Possibilistic C-Means results yield a value of 0.78821, with the best number of clusters being 4. The conclusion is that the Fuzzy C-Means method is better with optimal clusters is 4.

In Figures 2, 3, and 4 regarding cluster validity index results to provide conclusions, the criteria must meet the Partition Entropy Index values that are close to small or close to 0, as well as the Partition Coefficient Index and Modified Partition Coefficient Index values that are close to large or close to 1. Based on these criteria, the results of Fuzzy C-Means are better than all the experiments that have been carried out, with the values of the Partition Entropy Index at 0.21566, the Partition Coefficient Index at 0.88078, and the Modified Partition Coefficient Index at 0.82117.

As well as, cluster validity index results on the Fuzzy C-Means method from all experiments carrying out grouping or clustering in case studies of industrial business data that meet these criteria, the number of clusters is 3. The following are the results of members of the industrial business data cluster in Tanah Bumbu Regency in 2022 seasoning uses the Fuzzy C-Means method with 3 clusters.

Table 6. Cluster Results from Industrial Business Data

Cluster	*Number of Members*	*Cluster Labels*
Cluster 1	2483	Low Competitive Industry
Cluster 2	361	Medium Competitive Industry
Cluster 3	85	Highly Competitive Industry

In Table 6, it is concluded from the expert opinion of the Industry Office that cluster labels are based on the ability of the company or industry to face competitive challenges from its foreign competitors. Then the label helps analysts or decision makers to understand the condition of a particular industry and identify companies where improvements may be needed so that they can continue to grow in the industry. Obtained in cluster 1 has a total of 2483 data members. The characteristics in cluster 1 are obtained from data samples that have in common, namely the lowest membership for the Male Labor and Female Labor features and have low values for the Value of Investment, Production Value, and BB Value features. Investment, Production Value, and BW/BP Value features. Therefore, cluster 1 is considered as a group with the lowest level of economic activity compared to other clusters and is referred to as a low-competitive industry. Cluster 2 has 361 members. The characteristics in cluster 2 are obtained from data samples that have in common, namely medium membership for the Male Labor and Female Labor features and have medium values for the Investment Value, Production Value, and BB/BP Value features. Therefore, cluster 2 is considered a group with a moderate economic level compared to the previous group and is referred to as a medium competitive industry. And cluster 3 has a total of 85 members. The characteristics in cluster 3 obtained from the data sample that have in common are the highest membership for the Male Labor and Female Labor features and have a high average value for the Investment Value, Production Value, and BB / BP Value features. Therefore, cluster 3 is considered a group with a high economic level compared to the previous group and is referred to as a highly competitive industry.

Previous research shows that Fuzzy C-Means produces a Partition Entropy Index value of 0.50002 and a Partition Coefficient Index of 0.99998 on health facility data, as well as a Partition Coefficient Index value of 0.74749 and a Partition Entropy Index of 0.62123 on public health facility data. Another study with data on the distribution of contraceptive users showed that Fuzzy C-Means had the highest Modified

Partition Coefficient of 0.46407. However, this recent study used industrial business data and compared Fuzzy C-Means with Fuzzy Possibilistic C-Means. The results show that Fuzzy C-Means produces a Partition Entropy Index of 0.21566, Partition Coefficient Index of 0.88078, and Modified Partition Coefficient Index of 0.82117 in cluster 3. These results are better than previous studies, indicating an improvement in the quality of results due to the use of better datasets. This is a positive achievement that shows progress in understanding or modelling the phenomenon under study.

CONCLUSIONS AND RECOMMENDATIONS

The conclusions obtained from this study based on the results of the discussion with 5 experiments and the number of clusters obtained show that the Fuzzy C-Means method is the best performance method in case studies of industrial business data. The results of the cluster validity index based on the Partition Entropy Index is 0.21566, the Partition Coefficient Index is 0.88078, and the Modified Partition Coefficient Index is 0.82117. The best number of clusters is 3, with low-competitive industrial clusters, medium-competitive industrial clusters, and highly competitive industrial clusters. With this, it is hoped that further research will normalize data, apply other data types and attributes, and apply clustering and other evaluation methods.

REFERENCES

Amaliyah, L.R., Rahim, A., & Septiani, A. (2022). Comparison Of Fuzzy C-Means , Fuzzy Possibilistic C-Means and Possibilistic Fuzzy C-Means Algorithms on The Distribution Of Contraceptive Users In NTB Province. *The 2nd International Seminar of Science and Technology*.

Amirah, M. M. A., Widodo, A. W., & Dewi, C. (2017). Pengelompokan Lagu Berdasarkan Emosi Menggunakan Algoritma Fuzzy C-Means. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, *1*(12), 1526–1534.

Andani, S. R. (2013). Fuzzy Mamdani dalam Menentukan Tingkat Keberhasilan Dosen mengajar. *Seminar Nasional Informatika 2013*, *2013*(semnasIF), 57–65.

Apsari, G. R., Pradana, M. S., & Chandra, N. E. (2020). Implementasi Fuzzy C-Means dan Possibilistik C-Means Pada Data Performance Mahasiswa. *Unisda Journal of Mathematics and Computer Science (UJMC)*, *6*(2), 39–48. DOI: 10.52166/ujmc.v6i2.2392.

Badan Pusat Statistik. 2023. Industri Besar dan Sedang. Retrieved from http://www.bps.go.id/subject/9/industri-besar-dan-sedang.html.

Correa, C., Valero, C., Barreiro, P., Diago, M. P., & Tardaguila, J. 2011. A comparison of fuzzy clustering algorithms applied to feature extraction on vineyard. In *Proceedings of the XIV Conference of the Spanish Association for Artificial Intelligence* (pp. 7-11). https://oa.upm.es/9246

David, D., Lauro, M. D., & Herwindiati, D. E. (2020). Sistem Prediksi Customer Loyalty Dengan Metode RFM dan Fuzzy C-Means. *Computatio : Journal of Computer Science and Information Systems*, *4*(1), 33. DOI: 10.24912/computatio.v4i1.7099.

Firdaus, H. S., Nugraha, A. L., Sasmito, B., & Awaluddin, M. (2021). Perbandingan Metode Fuzzy C-Means Dan K-Means Untuk Pemetaan Daerah Rawan Kriminalitas Di Kota Semarang. *Elipsoida : Jurnal Geodesi dan Geomatika*, *4*(01), 58–64. DOI: 10.14710/halal.v%vi%i.9219.

Ganbold, G., & Chasia, & S. (2017). Comparison between Possibilistic c-Means (PCM) and Artificial Neural Network (ANN) Classification Algorithms in Land use/ Land cover Classification. *International Journal of Knowledge Content Development & Technology*, *7*(1), 57.

Grover, N. (2014). A Study of Various Fuzzy Clustering Algorithms. *International Journal of Engineering Research*, *3*(3), 177–181. DOI: 10.17950/ijer/v3s3/310.

Handoyo, S., Widodo, A., Nugroho, W. H., & Purwanto, I. N. (2019). The implementation of a hybrid fuzzy clustering on the public health facility data. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(6), 3549–3554. DOI: 10.30534/ijatcse/2019/135862019.

Haqiqi, B. N., & Kurniawan, R. 2015. Analisis Perbandingan Metode Fuzzy C-Means dan Subtractive Fuzzy C-Means. *Media Statistika*, *8*(2). DOI: 10.14710/medstat.8.2.59-67.

Herlinda, V., & Darwis, D. (2021). Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means. *Darwis, Dartono*, *2*(2), 94–99. DOI: 10.33365/jtsi.v2i2.890.

Mashfuufah, S., & Istiawan, D. (2018). Penerapan Partition Entropy Index, Partition Coefficient Index dan Xie Beni Index untuk Penentuan Jumlah Klaster Optimal pada Algoritma Fuzzy C-Means dalam

Pemetaan Tingkat Kesejahteraan Penduduk Jawa Tengah. *Proceeding of The URECOL*, 51–60.

Merliana, P.N., Ernawati, & Santoso, A. J. (2015). Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means, 978–979.

Nidyashofa, N., & Istiawan, D. (2017). Penerapan Algoritma Fuzzy C-Means untuk Pengelompokan Kabupaten / Kota di Jawa Tengah Berdasarkan Status Kesejahteraan Tahun 2015. *The 6th University Research Colloquium*, (September), 23–30.

Ozdemir, O., & Kaya, A. (2019). Comparison of FCM, PCM, FPCM and PFCM Algorithms in Clustering Methods. *Afyon Kocatepe University Journal of Sciences and Engineering*, *19*(1), 92–102. https://10.35414/akufemubid.429540

Putri, G. N. S., Ispriyanti, D., & Widiharih, T. (2022). Implementasi Algoritma Fuzzy C-Means Dan Fuzzy Possibilistics C-Means Untuk Klasterisasi Data Tweets Pada Akun Twitter Tokopedia. *Jurnal Gaussian*, *11*(1), 86–98. DOI: 10.14710/j.gauss.v11i1.33996

Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning dengan Teknik Supervised dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, *7*(2), 156. DOI: 10.51211/biict.v7i2.1422.

Rouza, E., & Fimawahib, L. (2020). Implementasi Fuzzy C-Means Clustering dalam Pengelompokan UKM Di Kabupaten Rokan Hulu. *Techno.Com*, *19*(4), 481–495. DOI: 10.33633/tc.v19i4.4101.

Somvanshi, M., & Chavan, P. (2016). A Review of Machine Learning Techniques Using Decision Tree and Support Vector Machine. *Integerernational Conference on Computing Communication Control and Automation (ICCUBEA)*, 1–7. DOI: 10.1109/ICCUBEA.2016.7860040.

Thilagaraj, T., & Sengottaiyan, N. (2019). Implementation of fuzzy C-means and fuzzy possibilistic C-means algorithms to find the low performers using R-tool. *International Journal of Scientific and Technology Research*, *8*(8), 1697–1701.

Widiyanto, M. T. A. C. (2019). Perrbandingan Validitas Fuzzy Clustering pada Fuzzy C - Means Dan Particle Swarms Optimazation (PSO) pada Pengelompokan Kelas. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *4*(1), 22. DOI: 10.14421/jiska.2019.41-03.