Optimization of MSMEs Clustering in Sampang District Using K-Medoids Method and Silhouette Coefficient Method

Muhammad Iqbal Firmansyah[1*], Yeni Kustiyahningsih[2], Eza Rahmanita[3], Mochammad Syahrul Abidin[4], Budi Dwi Satoto[5]

[1,2,3,4,5]Department of Information Systems, Universitas Trunojoyo Madura, Bangkalan, East Java, Indonesia
E-mail: [1*]iqbalcode.nt@gmail.com, [2]ykustiyahningsih@trunojoyo.ac.id, [3]eza.rahmanita@trunojoyo.ac.id, [4]syahrulabidin155@gmail.com, [5]budids@trunojoyo.ac.id

(Received: 31 Oct 2024, revised: 8 Nov 2024, accepted: 12 Nov 2024)

Abstract

Micro, Small, and Medium Enterprises (MSMEs) are an important sector in the economy, playing a significant role in creating jobs and driving local economic growth. This study aims to identify the business development patterns of MSMEs in Sampang District using the K-Medoids method. The background issue raised is the lack of appropriate segmentation for MSMEs, which complicates the efforts of the government and business actors in designing suitable development strategies. The dataset used consists of 1,276 MSME data points with six variables: Type of Business, Number of Workers, Production Capacity, Revenue, Assets, and Business License. The data processing steps include data conversion, one-hot encoding, and normalization to ensure uniformity. Clustering is performed using the Elbow method to determine the optimal number of clusters, with K=4 chosen as the optimal cluster number based on the highest Silhouette Coefficient value of 0.5662 compared to other K values. The Silhouette Coefficient values for K=2 are 0.3711, K=5 is 0.5389, K=7 is 0.5201, and K=9 is 0.4737. The clustering results show that this cluster encompasses various types of services, trade, to food and beverages sectors. This segmentation can support data-driven decision-making at the village level. Although this research shows promising results, it is recommended to expand the quantity and variety of data and consider external factors affecting MSME performance. Thus, this study makes a valuable contribution to understanding the business characteristics of MSMEs in Sampang District.

Keywords: K-Medoids, Elbow, Silhoutte Coefficient, Clustering, Similarity.

I. INTRODUCTION

MSMEs have an extremely strategic function in the development of the national economy [1]. Law Number 11 of 2020 on Job Creation has replaced Law Number 20 of 2008 and has given a much sharper legal basis in terms of assets and revenues from MSMEs for the development of this sector [2]. MSMEs also participate in the generation of national income and thus are very important for ensuring community welfare by providing job opportunities and driving economic growth with innovation. In this regard, support for MSMEs has been quite instrumental in achieving national competitiveness, apart from building a robust foundation for people-oriented economic development [3].

However, starting from early 2020, the COVID-19 pandemic sweeping the world has already hit the MSME sector very seriously. Due to mobility restrictions, market closure, and diminished consumer purchasing power, small businesses have witnessed extreme revenue decline. In addition to that, other complicating factors were difficulties in accessing raw materials, disruption in supply chains, and limited access to capital. As a result, many MSMEs have been compelled to temporarily close their businesses, while others closed down completely [4].

The backbone of the economy is always MSMEs during the crisis. In the post-pandemic era, support for MSMEs is very important to revive the national economy and increase the resilience of a people-centered economy. Innovation and digital transformation can be one of the solutions that allow MSMEs to adapt to new conditions and increased competitiveness in the digital era [5].

The MSME sector in Madura has grown significantly enough to contribute to the local economy [6]. The Sampang District can be said to be one of the districts that is getting quite significant in MSME development. In Sampang District, MSMEs have a bright future and an important role in creating employment opportunities and increasing community welfare [7]. However, they also face various challenges, such as limited access to capital, low-quality human resources, and difficulties in accessing market information that could support their business growth.

The K-Medoids technique of clustering was used in this study to classify MSMEs based on some identified characteristics. This algorithm separated the data into a number of groups based on similarities between various points of data. Among the advantages, K-Medoids can manage outlier-sensitive data and provide more stability in the output of clustering than other clustering methods. This can be useful in mapping MSMEs by type of business, capacity, and growth potential, and thus intervention strategies can be addressed to the needs of each group accordingly [8].

The elbow method is a common technique that optimizes the clustering process, determining the optimal number of clusters. It works by visualizing the total Sum of Squared Errors as a function of the number of clusters and finding that elbow point beyond which increasing the number of clusters does not substantially reduce the SSE. This is important in elbow optimization because it minimizes the loss of important information when the number of clusters chosen is too few or raises analysis complexity when there are too many without providing substantial benefits, as seen in [9].

Apart from the elbow method, there is another important tool to evaluate clustering quality-the Silhouette Coefficient. It gives a measure of how well a data point is positioned within its assigned cluster in comparison with other clusters. Scores range from -1 to 1. A score close to 1 means the data point is well-placed in its cluster. The score of about 0 stands for a data point being on or very close to the transition between two clusters. And finally, a negative score means that the data point probably is placed in the wrong cluster. Therefore, it may be that the Silhouette Coefficient will give more insight into the good quality of clustering being performed, and it could also serve to assist in decision-making regarding an appropriate number of clusters [10].

While the previous study has applied a K-Medoids method to cluster COVID-19 vaccination recipients in DKI Jakarta for targeting vaccine distribution, it reported an optimal cluster number of 6 since this showed the lowest DBI value of 0.209. Besides, clustering techniques play a significant role in strategic decision-making. This is similarly applied in clustering MSMEs based on the study of business characteristics [4]. A study entitled "Mapping the Characteristics of KIP-K Scholarship Recipients Using the K-Means++ Algorithm" utilized this very algorithm in clustering the students into various groups based on both their academic and non-academic attributes. The results of the mapping showed that, using a graph of the Elbow method, k=2. The highest silhouette coefficient was 0.7523 and the smallest DBI was 0.49053 [11]. Both stress the importance of clustering techniques in handling immense datasets in support of strategic decisions, hence giving a clue to apply to clustering MSMEs based on business characteristics.

While the research questions of this study focus on how to identify business development patterns with the K-Medoids method for MSMEs in the Sampang District, the differences in the evaluation method also set this study apart from any previous research. In the current study, the Silhouette Coefficient was used as the method of evaluation that stipulates how well certain data points fit into a defined cluster by measuring distances between different clusters. This is in contrast to prior studies, which utilized the Davies-Bouldin Index as a way of measuring clustering effectiveness. The Silhouette Coefficient provides a more direct look at cluster separation quality, with an optimal value more relevant to clustering needs by MSMEs in terms of similarity in business characteristics.

Based on this consideration, this study recommends clustering MSMEs in Sampang District using the K-Medoids method. The K-Medoids algorithm is chosen in this regard because of its effectiveness in data sensitive to outliers and greater stability in clustering processes compared to other methods in previous research. Although K-Medoids has been widely applied for MSME clustering, this research underlines the fact that not too many have dealt with the challenges of clustering MSMEs in the Sampang District regarding clustering not only based on business characteristics but also the distribution of MSME actors in each village. The purpose of this research is to find a pattern in MSME development in Sampang District using the K-Medoids method so that the segmentation of MSMEs is more specific and focused. This will be helpful in planning strategies for developing MSMEs by the government and business actors who might fit better with the needs and potentials of each MSME group and support wiser evidence-based decisions on village development.

II. RESEARCH METHODS

The method that will be used in this study is the K-Medoids. Figure 1 is a conceptual model of this research, while Figure 2 is the flow of this study.

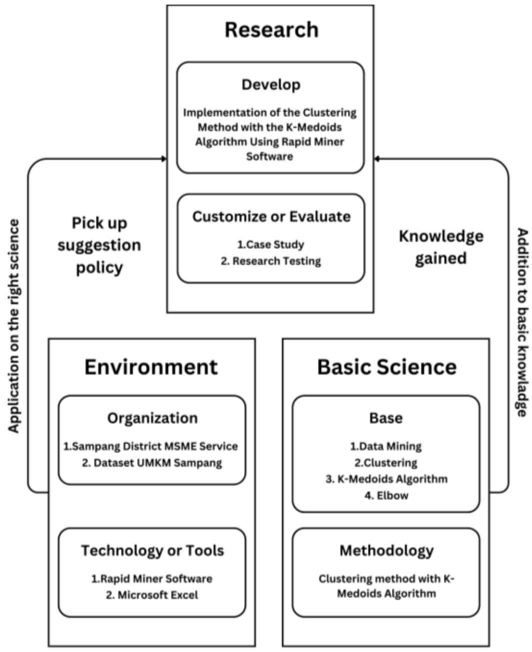

Figure 1. Conceptual Model

ISSN 2549-8037, EISSN 2549-8045
DOI: 10.34148/teknika.v14i1.1116

Based on Figure 1, it concludes that the above modeling is based on the case study of MSMEs in Sampang District. This modeling is expected to determine MSME clustering so that it will be easier for the government to consider what MSMEs need appropriate socialization or guidance. The actors in this community are local residents and the Sampang District government, who work together to build an increasingly conducive ecosystem for growing MSMEs. Data mining with RapidMiner software is used in this study; therefore, effective and efficient data analysis can be performed. With the result of this analysis, the Sampang District government can make a better decision by using the generated MSME clustering so that it allows special attention to those MSMEs that need extra support. Moreover, the description in the following section depicts how this study was conducted by presenting its flow with the keysteps about data collection, processing, and analysis.

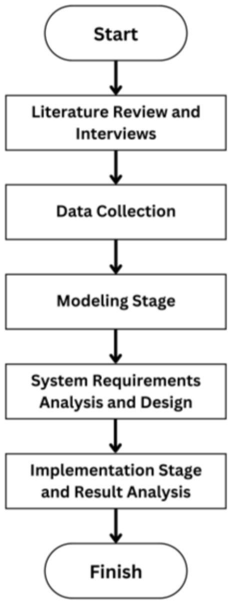

Figure 2. Research Flow

A. Research Stages

The stages of the research can be seen in Figure 2, which consists of:

1. Literature Study and Interviews
   In this stage, literature reviews and journal studies are carried out referencing theories that underlie the process of data mining, clustering, data pre-processing, the K-Medoids method, one-hot encoding, and the Elbow method to find the number of clusters. In addition, it also performs interviews and requests for data with the Cooperative and Micro Business Office of Sampang District.
2. Data Collection
   The data collection stage includes:
   - Data on MSMEs in Sampang District contains 1,276 records.
   - Data on MSMEs are used as research objects available in Sampang District.
   - Information Data taken from the Cooperative and Micro Business Office of Sampang District.
   - Data indicators from both the office and MSME actors are consistent with the data in Table 1.
3. Modeling Stage
   In the process of modeling in the MSME mapping application with clustering-based approach, the criteria used in this research are adjusted with the needs of MSMEs in Sampang District. Pre-processing is the next process after cluster modeling; it means converting data into numerical format using the one-hot encoding method so that each categorical variable can be analyzed in quantitative manners. This method converts categorical data into numerical data in 0 and 1 forms. The Elbow method will be applied before clustering to find out the best number of clusters. It will evaluate the sum of squared errors for various numbers of clusters until it reaches the elbow point, which shows the best number of clusters. The K-Medoids Clustering starts with the normalization of data followed by centroid determination after the determination of the optimal number of clusters. Euclidean Distance is calculated for finding the distance between different data points, and the clustering proceeds by updating the cluster in an iterative manner until stable results are obtained. After the dataset forms into relatively stable cluster, further analysis of the result is conducted to provide insight applicable for MSMEs development in Sampang District.
4. Needs Analysis and System Design
   This stage involves the specification of hardware and software requirements for clustering Batik MSMEs. The hardware specifications are as follows:
   - Intel Core i5 Processor
   - 8 GB RAM
   - 1 TB Hard Disk

   The software specifications include:
   - Windows 10 Operating System
   - RapidMiner
   - Microsoft Excel
   - StarUML

   In the development of this software, the following outputs would be generated:
   - Database information on MSMEs obtained from the Cooperative and Micro Business Office of Sampang, Madura.
   - Recommen dations and clustering of MSMEs based on the K-Medoids Clustering method.
   - Identifying the weights of variables that most influence clustering.
   - Testing the implementation for analysis to guarantee proper functioning and ensure the validity of the results in clustering.
5. Implementation and Result Analysis Stage
   Implementation of the system model will be carried out in harmony with design and analysis of the system requirements that have been prepared beforehand. The result of the implementation of clustering will allow the MSME parties and the office to ascertain recommendations for improvement on each relevant indicator to guide efforts toward improving the productivity of MSMEs in Sampang

DOI: 10.34148/teknika.v14i1.1116

ISSN 2549-8037, EISSN 2549-8045

District. The preprocessing will consist of the stages in the trials, to which one-hot encoding will be applied so that each categorical variable can be analyzed quantitatively. Using this method, the categorical data was changed into numerical data with the form 0 and 1. Next, the determination of the optimal number of clusters using the Elbow method and centroid determination will be done. These data points use a method called Euclidean Distance to find the distance between them. After the clustering is done, the cluster results will be optimized and evaluated using Silhouette Coefficient to ensure that the quality of the separation created by the cluster is good.

B. Integration of K-Medoids and Elbow Method

The integration of both methods can be seen in the Figure 3 below:

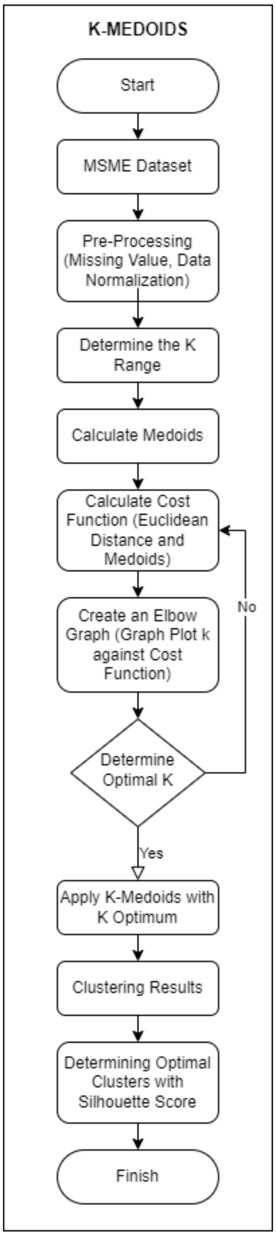

Figure 3. Integration of K-Medoids and Elbow

Figure 3 describes from the very beginning-which in this case is collecting data to be analyzed. Prior to the clustering, the data will undergo pre-processing to guarantee that the data is in good condition, including handling missing values and normalization of the data. After all the preprocessing is done, the data will be grouped by utilizing the K-Medoids algorithm. The Elbow method will be used to determine the best number of clusters, including a graphical representation provided for this purpose. In addition, the Silhouette Coefficient gives information on the separation of each data point from other clusters in quality clustering evaluation.

C. K-Medoids Clustering

K-Medoids is a clustering algorithm that aims at the segregation of data by their medoids, which are the most representative data points in a cluster. The steps of the K-Medoids algorithm are as follows:

1. Choose K medoids from the dataset.
2. Calculate the distance between each object x_i and each medoid y_j.
3. Assign each object x_i to the nearest medoid.
4. The new positions of medoids could be obtained from the Formula 1:

$$y_j = arg \min_{x_i \epsilon C_j} \sum_{x_k \epsilon C_j} d(x_i, x_k) \quad (1)$$

where C_j is the cluster assigned to the medoid y_j and $d(x_i, x_k)$ is the distance between objects x_i and x_k.

5. Repeat Steps 3 and 4 until convergence of clusters, that is no change in cluster assignments.

D. Elbow Method

The elbow method is used to determine the ideal number of clusters in cluster analysis. In that context, it calculates the Sum of Squared Errors with respect to different values of K-the number of clusters-and then plots them as a graph. Following is the Formula 2 for calculating the SSE for any value of K:

$$SSE(K) = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \left\| x_j^{(i)} - c_i \right\|^2 \quad (2)$$

Where:

- K is the number of clusters being evaluated.
- $x_j^{(i)}$ is the j-th data point in the i-th cluster.
- c_i is the center (centroid) of the i-th cluster.
- n_i is the number of data points in the i-th cluster.
- $\left\| x_j^{(i)} - c_i \right\|^2$ is the squared distance between the data point $x_j^{(i)}$ and the center of the cluster c_i.

E. Silhouette Coefficient

Silhouette Coefficient is a measure to evaluate the quality of clustering. The value range of the Silhouette Coefficient is from -1 to 1, where a higher value indicates good clustering. The Silhouette Coefficient is computed in the following steps:

ISSN 2549-8037, EISSN 2549-8045
DOI: 10.34148/teknika.v14i1.1116

1. Calculate the distance between each object x_i in cluster C_k and other objects.
2. For each object x_i, compute the Silhouette Coefficient using the following Formula 3:

$$s(x_i) = \frac{b(x_i) - a(x_i)}{\max(a(x_i), b(x_i))} \quad (3)$$

Where:
- $a(x_i)$ the mean distance of object xi to all other objects in the same cluster.
- $b(x_i)$ the mean distance of object xi to the nearest object in a different cluster.

3. The average of all Silhouette Coefficient values gives an overall clustering score.

III. RESULTS AND DISCUSSION

A. Dataset Processing

In this case, the process commences with dataset processing. Processed dataset consists of 6 variables: Business Type, Number of Workers, Production Capacity, Revenue, Asset, and Business License that make up a total of 1,276 MSME data originating from Sampang District listed in Table 1.

Table 1. Transformation of UMKM Cluster Criteria

Criteria	Values	Data Conversion
Number of Employees	7 - 8	4
	5 - 6	3
	3 - 4	2
	1 - 2	1
Production Capacity	3,751 – 5,000	4
	2,501 – 3,750	3
	1,251 – 2,500	2
	<= 1,250	1
Revenue	Rp. 563,187,501 - Rp. 750,000,000	4
	Rp. 375,575,001 - Rp. 563,187,500	3
	Rp. 187,962,501 - Rp. 375,575,000	2
	<= Rp. 187,962,500	1
Assets	Rp. 93,762,501 - Rp. 125,000,000	4
	Rp. 62,525,001 - Rp. 93,762,500	3
	Rp. 31,287,501 - Rp. 62,525,000	2
	<= Rp. 31,287,500	1
Business License	Exists	1
	Does Not Exist	0

B. Pre-Processing Stage

The next step after processing the dataset is the transformation of MSME data, as shown in Table 2. To do clustering, the data must first undergo a pre-processing step using one-hot encoding, as shown in Table 3. Table 4 shows the results of the normalization step, which was done to simplify the data.

Table 2. UMKM Dataset of Sampang District

No	Employe	Capacity	Revenue	Assets	License
1	2	3	1	1	1
2	1	1	2	2	0
…	…	…	…	…	…
1,275	2	2	1	2	1
1,276	4	1	1	2	0

For example, in Table 2, Sampang District MSME dataset with transformed data from the conversion reference obtained from Table 1. Table 1: Conversion basis for the value of each variable to numerical values. Clearly spelled out in this table are variables like the number of employees, production capacity, revenue, and assets. These are then stratified into specified ranges where a numerical value represents a group. For example, the variable of the number of workers is transformed into numerical values by groups. That is, 7-8 workers are transformed into 4, and 1-2 workers are transformed into 1. It is done to enhance efficiency and coherence in the whole analysis process and further on at the stage of data clustering. Another variable is the business license status, which has one-hot encoding: 1 for 'Has Business License' and 0 for 'Does Not Have Business License', as depicted in Table 3.

Table 3. One-Hot Encoding Stage

No	Business Name	Business License
1	Durroh Konveksi	1
2	Budidaya LOBSTER Farhan Abdillah	0
…	…	…
1,275	UD.Syaiful Jaya	1
1,276	Jamu Tradisional Siti Fatima	0

Table 3 shows the result from one-hot encoding of the variable Business License. Each row in this table represents the MSMEs-'Durroh Konveksi' and 'UD.Syaiful Jaya'. This variable was encoded in binary format, with 0 meaning without a business license and 1 meaning with a business license. All these one-hot encoding steps are really preparing the data for analysis, furthering the clustering model performance.

Table 4. Data Normalization Stage

No	…	…	Revenue	Assets	Business License
1	…	…	0.1436	1.0000	1.0000
2	…	…	0.1940	0.2086	0.0000
…	…	…	…	…	…
1,275	…	…	0.0482	0.2086	1.0000
1,276	…	…	0.0327	0.2086	0.0000

Table 4 presents the results of data normalization, including some variables such as Revenue, Assets, Business License. Data normalization is a process of scaling numerical values in a given range-usually between 0 and 1-for easier analysis and modeling. Each line here is an MSME, and the values here, for instance, 0.0529 for Revenue and 0.0647 for Assets, are results obtained from normalization. Normalization may reduce bias that might arise because of differences in scale among variables, thereby increasing the accuracy of data analysis and clustering. Values in Business License column are left as binary, whether a business license exists or not; 0 is no license, and 1 is there is a license.

C. Modeling and Evaluation

After the dataset pre-processing stage, randomly decide on the cluster centroids and then calculate the distance to the nearest cluster using the Euclidean Distance. Several trials in the clustering process are done with the values of K, which ranges from 2 to 10. From these several trials, the value of optimum K is decided by Elbow method. The result of Elbow calculations for each K value can be seen in Figure 4.

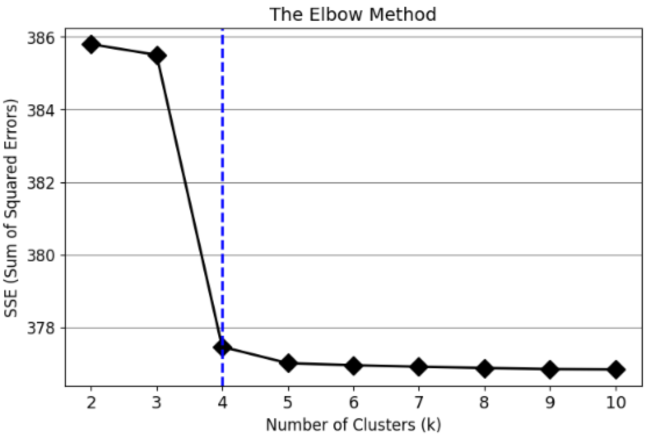

Figure 4. Elbow Graph

Figure 4, The graph in this figure depicts the value of K = 4. This forms the most marked elbow point compared to other values of K within the range of K=2 to K=10. That means the elbow method used classifies K = 4 as the best choice for the number of clusters in this dataset. After establishing the value of the optimal K, the next step is the calculation of the Silhouette Coefficient for K=4 in comparison to a variety of clustering results by using K value from 2 to 10. The comparison of the Silhouette Coefficient values between K=2 and K=10 are depicted in Table 5.

Table 5. Silhouette Score Values

No	Values K	Values Silhouette Coefficient
1	K=2	0.3711
2	K=3	0.0819
3	K=4	0.5662
4	K=5	0.5389
5	K=6	0.5165
6	K=7	0.5201
7	K=8	0.5021
8	K=9	0.4737
9	K=10	0.4701

Table 5, Comparative results of the Silhouette Coefficient values for varied K values. It follows from the table that Silhouette Coefficient reaches its highest value for K=4 with the score 0.5662 close to 1, reflecting the most optimal quality of clustering. Concretely speaking, for K=4, there is more homogeneity between clusters compared with other K, and data is much more salient between clusters. Based on this study, therefore, K=4 was chosen as the most appropriate number of clusters for clustering MSMEs. The distribution of MSMEs in each cluster for K=4 is presented in Table 6.

Table 6. Results of Cluster 4

Business Name	...	Capacity	Revenue	…	Cluster
Durroh Konveksi	...	0.0164	0.0395	...	C4
Kholifah Taylor	...	0.0245	0.0195	...	C4
Maria Tailor	...	0.0787	0.0195	...	C4
Muafana Taylor	...	0.0843	0.0195	...	C4
UD Sa'i	...	0.0946	0.0328	...	C4
...	...	...	...	...	...
...	...	...	...	...	...
Toko Eva	...	0.0282	0.0015	...	C4
UD. Lukman	...	0.1996	0.1996	...	C4
Toko B. Rus	...	0.0873	0.0062	...	C4
Nasi Bu Hai	...	0.0799	0.0262	...	C4
Dapoer Haidar	...	0.0952	0.0328	...	C4
Kripik Mbak Harimah	...	0.0120	0.2663	...	C4

Table 6 presents data of MSMEs in the clusters. From the list of MSMEs, it is observed that the clusters entail a wide range of services, trade, to food and beverages. This diversity of industries somehow informs that the cluster with K=4 includes MSMEs with different characteristics and possibly similar types of business, number of employees, production capacity, turnover, assets, and business licenses that form the basis for clustering in this analysis.

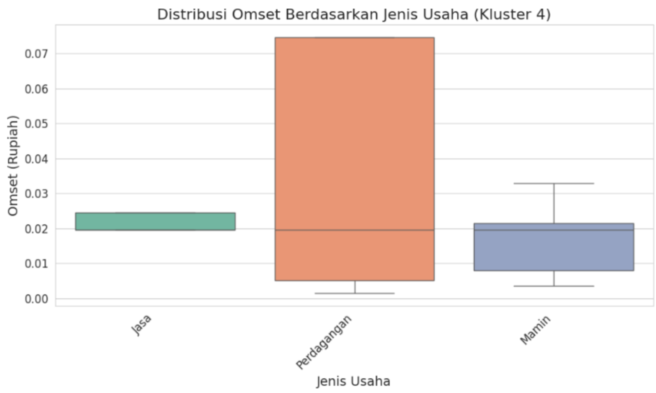

Figure 5. Distribution of Revenue Graph

Figure 5 depicts the graph of the average revenue distribution of each business type in the fourth cluster of MSME clustering in Kecamatan Sampang using K-Medoids. From the graph, it is observable that the average of revenue from an trade business is higher compared with the other two business types in this cluster, after which follows the services sector. While the food and beverages sectors are relatively lower in average revenue. These clusters are determined for the reason that it will find out the pattern in the business development about MSME, based on the characteristics of business types and its revenue generated.

IV. CONCLUSION

This study successfully mapped business development patterns of MSMEs in Sampang District with the K-Medoids method, which was used to group MSMEs by their business characteristics so that segmentation can be done more precisely. It is expected that this shall assist the government and business actors in formulating strategies that can adjust to

DOI: 10.34148/teknika.v14i1.1116

the needs of each group and assure data-driven decision-making at the level of the village. The results and discussion in the next section identify the average distribution of revenues among various types of businesses in Cluster 4, which indicates that, among others, businesses in trade are among the ones that have the highest average revenue, followed by the services sectors, while the food and beverages sectors show relatively lower average revenues. Such a clustering allows the analysis of MSME business development patterns based on the type of business and the revenue generated, providing strategic insight to determine the support and development of MSMEs.

Business types, number of employees, production capacity, revenues and assets, number of business licenses are included in the used dataset. Such data preprocessing as data conversion, one-hot encoding, and normalization were performed to create a harmonized dataset that is easy to analyze. The best number of clusters was determined by running clustering using the Elbow method. This indicated that the preferable number of clusters is K = 4, for which the Silhouette Coefficient is highest, approximately 0.5662, as compared to other values of K. Other values of K include K = 2 with a Silhouette Coefficient of approximately 0.3711, K = 5 with approximately 0.5389, K = 7 with approximately 0.5201, and K = 9 with approximately 0.4737.

Herein, the method for the evaluation of clustering uses the Silhouette Coefficient, which is different from earlier works using DBI. The Silhouette coefficient is a more direct measure of quality regarding inter-cluster separation, which estimates an average of inter-cluster and intra-cluster distances of samples; thus, it has helped to obtain optimal clustering based on similarity in business features among MSMEs. This method offers a more relevant optimal value toward the needs of clustering MSME, thus generating stronger segmentation to represent variation in business characteristics among MSME groups within Sampang District.

The evaluation of clustering can be done by showing that the SSE value decreases significantly at K=4. This K=4 cluster consists of the types of MSMEs from services, trade, to food and beverages sectors. Each cluster represents a group of MSMEs with similar business characteristics; that is, business type, number of employees, production capacity, revenue, assets, and business licenses. Segmentation thus provides a sound basis for more focused development strategies, supporting evidence-based planning by local governments and business actors.

The first advantage of the K-Medoids method used in this study is that it is capable of handling asymmetric data, reducing the influence of outliers, which are quite often present in MSME data due to differences in size and their own particular business characteristics. Moreover, with the Silhouette Coefficient as the metric of evaluation, the result of clustering is more measurable in terms of quality regarding the separation of clusters.

However, there are a number of limitations that have to be considered in view of the positive clustering result. First is the limited amount of data used, which might affect the generalization of research findings. Furthermore, this study does not account for any external factors that may affect the performance of MSMEs in general, such as government policies and market conditions, or broader socio-economic factors. Ambiguity in the measurement for some variables, such as production capacity and revenues, might also blur accurate clustering.

Further studies are recommended with increased quantity and diversity to improve the result of this analysis. In future studies, external variables affecting MSMEs also need to be taken into consideration and to make more accurate measurements for the existing variables. Such a combination approach could be extended with other methods, like regression analysis or machine learning, to extend knowledge about the factors affecting MSME development in Sampang District.

REFERENCES

[1] D. Azmi Fadhilah and T. Pratiwi, "Strategi Pemasaran Produk UMKM Melalui Penerapan Digital Marketing," *Coopetition J. Ilm. Manaj.*, vol. 12, no. 1, pp. 17–22, 2021, doi: 10.32670/coopetition.v12i1.279.

[2] D. K. Harjono, *Tinjauan Terhadap Undang-Undang Nomor 11 Tahun 2020 Tentang Cipta Kerja*. 2021.

[3] S. Setiawansyah, A. T. Priandika, B. Ulum, A. D. Putra, and D. A. Megawaty, "UMKM Class Determination Support System Using Profile Matching," *Bull. Informatics Data Sci.*, vol. 1, no. 2, p. 46, 2022, doi: 10.61944/bids.v1i2.37.

[4] F. M. Trisuciana, D. Witarsyah, E. Sutoyo, and J. M. Machado, "Clustering of COVID-19 Vaccination Recipients in DKI Jakarta Using the K-Medoids Algorithm," *Proc. - Int. Conf. Adv. Data Sci. E-Learning Inf. Syst. ICADEIS 2022*, pp. 4–10, 2022, doi: 10.1109/ICADEIS56544.2022.10037509.

[5] S. Setioko, Y. Fitriani, and K. Munawaroh, "Strategi Peningkatan Usaha Kecil Mikro dan Menengah (UMKM) di Era Pandemi Covid-19 Pada Kota Metro," *J. Community Dev.*, vol. 2, no. 2, pp. 60–65, 2021, doi: 10.47134/comdev.v2i2.24.

[6] M. R. Adiyanto and E. Amaniyah, "Tingkat Kesadaran Sertifikat Halal Pelaku UMK di Pulau Madura," *AKSES J. Ekon. dan Bisnis*, vol. 18, no. 2, pp. 94–101, 2023, doi: 10.31942/akses.v18i2.10123.

[7] Y. M. Riszinin and T. R. Dwi Adi Nugroho, "Preferensi Konsumen terhadap Pembelian Keripik Singkong di UD. Sumber Mutiara Kecamatan Sampang Kabupaten Sampang," *Agriscience*, vol. 3, no. 1, pp. 58–71, 2022, doi: 10.21107/agriscience.v3i1.15209.

[8] A. S. Sunge, Y. Heryadi, Y. Religia, and Lukas, "Comparison of Distance Function to Performance of K-Medoids Algorithm for Clustering," *Proceeding - ICoSTA 2020 2020 Int. Conf. Smart Technol. Appl. Empower. Ind. IoT by Implement. Green Technol. Sustain. Dev.*, 2020, doi: 10.1109/ICoSTA48221.2020.1570615793.

[9] M. Lee, S. Lee, J. Park, and S. Seo, "Clustering and characterization of the lactation curves of dairy cows using K-medoids clustering algorithm," *Animals*, vol.

DOI: 10.34148/teknika.v14i1.1116

ISSN 2549-8037, EISSN 2549-8045

10, no. 8, pp. 1–14, 2020, doi: 10.3390/ani10081348.

[10] Y. Kustiyahningsih, A. Khozaimi, and J. Purnama, "Pengelompokan UMKM Batik Madura Menggunakan Metode K-Means dan Sillhoutte Coefficient," *Teknika*, vol. 13, no. 2, pp. 192–198, 2024, doi: 10.34148/teknika.v13i2.779.

[11] F. Nuraeni, D. Kurniadi, and G. Fauzian Dermawan, "Pemetaan Karakteristik Mahasiswa Penerima Kartu Indonesia Pintar Kuliah (KIP-K) menggunakan Algoritma K-Means++," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 437–443, 2023, doi: 10.32736/sisfokom.v11i3.1439.

DOI: 10.34148/teknika.v14i1.1116