JINAV: Journal of Information and Visualization ISSN: 2746-1440 (online)
Vol. 7 No. 1 (2026)
https://doi.org/10.35877/454RI.jinav4844

Clustering MSMEs in Jambi Province Using K-Means Algorithm for Policy Strengthening

Hetty Rohayani[a,*], Noneng Marthiawati[b], & Saleh Yaakub[b]

[a]Informatics Study Program, Faculty of Science and Technology, Muhammadiyah University of Jambi, Jambi, Indonesia
[b]Information Systems Study Program, Faculty of Science and Technology, Muhammadiyah University of Jambi, Jambi, Indonesia

Abstract

Micro, Small, and Medium Enterprises (MSMEs) play an important role in supporting economic growth in Jambi Province. However, the distribution of MSMEs across regencies and cities has different characteristics, requiring further analysis to identify distribution patterns. This study aims to cluster MSME distribution in Jambi Province using the K-Means algorithm as an unsupervised machine learning approach. The study used MSME data from 2019–2023 obtained from the Regional Office of Cooperatives and Small Enterprises of Jambi. Data processing was conducted using Python through preprocessing, the Elbow Method, K-Means clustering, and cluster evaluation using the Silhouette Coefficient and Davies-Bouldin Index (DBI). The results showed that the optimal number of clusters was three: Low Cluster (C0), Medium Cluster (C1), and High Cluster (C2). The Low Cluster included Merangin, Sarolangun, and Tebo Regencies, while the High Cluster included Batanghari, Kerinci, Muaro Jambi Regencies, and Jambi City. The evaluation results produced a Silhouette Coefficient value of 0.81 and a DBI value of 0.21, indicating that the clustering model has good quality and optimal cluster separation. This study can support regional development planning and targeted MSME policy formulation in Jambi Province.

Keywords: MSMEs, Jambi Province, K-Meams Algorithm, Clustering, Policy Strengthening

Received: 17 February 2026 Revised: 13 April 2026 Published: 30 April 2026

1. Introduction*

Micro, Small, and Medium Enterprises (MSMEs) play a significant role in supporting economic growth, creating employment opportunities, and improving community welfare. In Indonesia, MSMEs are recognized as one of the main contributors to both regional and national economic development(Ezeudo, M.T. and Biose, 2024; Maksum et al., 2020). In Jambi Province, the number of MSMEs continues to increase annually; however, their distribution across regencies and cities demonstrates diverse characteristics. These differences are influenced by various factors, including assets, turnover, business scale, and regional economic conditions. Such diversity highlights the importance of conducting data-driven analyses to better understand MSME distribution patterns and support more effective and targeted policy formulation.

The rapid development of Artificial Intelligence (AI), Machine Learning (ML), and data mining technologies has significantly transformed business analysis and decision-making processes(Bharadiya, 2023). AI technologies are increasingly utilized to improve efficiency, decision-making accuracy, and strategic innovation, particularly within small and medium enterprises(Kim & Seo, 2023). In the MSME sector, the integration of AI and data analytics enables the identification of hidden patterns, business segmentation, and enterprise characteristics more effectively than conventional descriptive approaches(Pamungkas et al., 2023; Ragazou et al., 2023). Consequently, AI-driven approaches have become increasingly relevant in supporting business intelligence, sustainability, and policy development for MSMEs.

* Corresponding author.
E-mail address: hettyrohayani@gmail.com

One of the most widely applied techniques in AI-based data analysis is clustering, an unsupervised machine learning method that groups data according to similarity characteristics to form homogeneous clusters(Syakur et al., 2018). Among clustering algorithms, K-Means is considered one of the most popular and effective methods because of its simplicity, computational efficiency, and strong capability in handling large datasets(Yuan & Yang, 2019). The performance of K-Means is highly dependent on determining the optimal number of clusters. To address this issue, Syakur et al., (2018) integrated K-Means with the Elbow Method to improve clustering effectiveness, while Viloria & Lezama, (2019) emphasized that proper cluster number determination significantly enhances clustering performance in SME-related databases.

Previous studies have widely implemented K-Means clustering in MSME-related research. Nugraha et al., (2021) applied cluster analysis to determine business strategies for MSMEs in Yogyakarta, while Nouvel & Dwi Mulyanto, (2025) clustered MSMEs based on assets and turnover. Similarly, Sutramiani et al., (2024) compared the performance of K-Means and DBSCAN for MSME grouping and found that K-Means produced competitive results. Other studies integrated K-Means with decision-support approaches, such as combining K-Means and AHP for MSME recommendation systems (Kustiyahningsih et al., 2024) and integrating K-Means with Apriori algorithms for digital marketing strategies in MSMEs(Maulidya et al., 2025). These findings demonstrate that clustering methods can effectively support business segmentation, recommendation systems, and strategic decision-making for MSMEs.

Research in MSME clustering has also evolved into more advanced approaches, including fuzzy clustering and hybrid machine learning models. Caraka et al., (2021) proposed optimized fuzzy geodemographic clustering to analyze MSME business vulnerability in Indonesia, while Atemoagbo et al., (2024) utilized fuzzy c-means clustering combined with t-SNE visualization to analyze MSMEs in Nigeria. Reyes-Ruiz & Hernandez-Hernandez, (2021) further introduced fuzzy clustering for determining MSME business size based on financial information. Beyond clustering, machine learning approaches have also been applied for MSME performance prediction, credit risk quantification, and strategic policy analysis(Boubaker et al., 2025; Cao, 2025). Stevenson et al., (2021) additionally demonstrated the effectiveness of deep learning approaches for small business default prediction. These studies collectively indicate that AI and machine learning technologies have substantial potential to strengthen MSME development and policy planning.

Despite the growing number of studies related to MSME clustering, research specifically focusing on MSME distribution patterns across regions in Jambi Province remains limited, particularly studies utilizing AI/ML approaches through the K-Means algorithm. Furthermore, most previous studies primarily focused on business segmentation, financial profiling, technology readiness, or marketing strategies, whereas regional distribution analysis for policy strengthening has received less attention. Therefore, this study addresses this gap by applying the K-Means clustering algorithm to analyze MSME distribution patterns across regencies and cities in Jambi Province using MSME data from 2019–2023.

This study applies the K-Means algorithm integrated with the Elbow Method to determine the optimal number of clusters and evaluates clustering quality using the Silhouette Coefficient and Davies-Bouldin Index (DBI). The findings are expected to provide a clearer understanding of MSME distribution patterns in Jambi Province and contribute to regional development planning and more targeted MSME policy strengthening.

2. Methods

This research employs a data mining method with a clustering approach using the K-Means algorithm to classify Micro, Small, and Medium Enterprises (MSMEs) in Jambi Province based on their business characteristics. This method was selected because it is capable of effectively grouping large amounts of data according to the similarity of attributes among the data. In addition, the K-Means algorithm offers an efficient computational process, is easy to implement, and is able to generate clear and well-structured clusters, making it suitable for MSME data classification(Ramadhan et al., 2025).

To ensure that the research process is conducted systematically, effectively, and in accordance with the research objectives, a research framework was designed to illustrate the overall problem-solving process. The research stages include data selection, preprocessing, data transformation, the clustering process using the K-Means algorithm, and the evaluation of clustering results. The overall research stages are illustrated in Figure 1.

Figure 1. Illustration of Research Stages

The figure 1 illustrates the stages of the data mining process using the K-Means Clustering algorithm. The process begins with:

2.1. Data Selection

At the data selection stage, the process of identifying, selecting, and collecting relevant data from the available operational data was carried out. The research data were obtained from the Regional Office of Cooperatives and Small Enterprises of Jambi through direct observation at the related institution. The data used consisted of the number of small, medium, and large industrial units in Jambi by regency/city during the 2019–2023 period. Data selection is an essential stage in data mining because it determines the relevance and quality of the dataset used for clustering analysis(Pamungkas et al., 2023).

2.2. Preprocessing

At the data preprocessing stage, data cleaning and preparation were carried out before applying the K-Means algorithm. This stage aimed to improve data quality in order to produce more accurate clustering results. The preprocessing process began with data completeness checking (data cleaning) to identify missing, inconsistent, or duplicate data. Incomplete data were then corrected or removed according to the research requirements. Furthermore, outlier detection was conducted to minimize the influence of extreme values on the clustering results.

After the cleaning process was completed, the data were selected based on relevant attributes, namely the number of small, medium, and large business units in each regency/city in Jambi during the 2019–2023 period. The data were then organized into a tabular format and ensured to be in numerical form so that they could be processed by the K-Means algorithm. The preprocessing stage resulted in a clean, consistent, and well-prepared dataset ready to be used in the transformation and clustering processes(Kondoj et al., 2025).

2.3. Transformation

At the data transformation stage, the MSME data that had been obtained were processed and transformed into a format suitable for the clustering process using the K-Means algorithm. The transformation process was carried out by organizing the data based on the selected attributes, namely the number of small, medium, and large business units in each regency/city in Jambi during the 2019–2023 period.

Furthermore, a data normalization process was conducted to minimize differences in scale among attributes so that each variable contributed proportionally to the clustering process. This stage also included data encoding, final dataset preparation, and conversion of the data into a numerical format that could be processed by the K-Means algorithm(Wibowo et al., 2023). The output of the transformation stage was a clean and structured dataset ready to be used in the clustering process(Lestari et al., 2025).

2.4. K-Means Clustering Process

The K-Means algorithm is a widely used data mining technique for clustering data by assigning each data point to the nearest cluster center (centroid), with the objective of maximizing similarity within clusters and minimizing similarity between clusters. Prior to applying the K-Means algorithm, data pre-processing was conducted to ensure that all attributes were on a comparable scale and to prevent features with larger numeric ranges from dominating the clustering results(Murdiono et al., 2026). In this study, Min–Max normalization was employed to transform the data into a uniform range. The normalization process is defined by the following equation:

Normalized Value = (((data – data min) * (new max – new min) / (data max - data min)) + new min)

This normalization step improves the stability of distance-based calculations and enhances the quality of the clustering results by enabling a more balanced and meaningful grouping of data points. Clustering using the K-Means method In generally done with the following algorithm. The steps of the K-Means Algorithm are as follows:

a. Determine the value of k or the number of Clusters on the data set

b. Determining the center value (Centroid). Determination of the centroid values at the initial stage is done randomly, while in the iteration stage the formula as below is used:

$$V_{ij} = \frac{1}{N_i}\sum_{k=1}^{N_i} X_{kj}$$

Description:

V_{ij} = The average centroid of the i-th cluster for the j-th variable.
N_i = Number of members of the i-th cluster.
i,k = Index of the cluster.
j = Index of variables.
X_{kj} = the k-th data value of the to j-th variable for that cluster.

c. Calculate the distance between the centroid point and the point of each object using Euclidean Distance. Euclidean Distance is an ordinary straight line distance between two points in Euclidean space, with the formula as below:

$$d_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Description:

d_e = Euclidean Distance.
i = Number of objects.
x,y = Object Coordinate point.
s,t = Centroid coordinate point.

d. Group the objects based on the distance to the nearest Centroid 5. Repeat steps 2 to 4, iterate until the Centroid is optimal.

2.5. Evaluation

The evaluation stage was conducted to assess the quality of the clustering results generated by the proposed model. In this study, clustering evaluation was performed using two methods: the Davies–Bouldin Index (DBI) and the Silhouette Coefficient(Terttiaavini, 2024).

The Davies–Bouldin Index (DBI) was used to evaluate cluster quality based on the distance among data points within the same cluster and the separation between clusters. This method aims to produce clusters with low intra-cluster distance and high inter-cluster distance. A lower DBI value indicates better clustering quality. Meanwhile, the Silhouette Coefficient was used to measure how well each data point fits within its assigned cluster compared to other clusters. A higher Silhouette Coefficient value indicates better clustering performance.

3. Result and Discussion

This section presents the results of MSME clustering in Jambi Province using the K-Means algorithm. The clustering results were analyzed to identify patterns and characteristics of MSMEs in each cluster, which can support policy strengthening and decision-making.

3.1. Data Selection

This chapter discusses the research findings on the use of the K-Means Clustering algorithm to facilitate the grouping of MSMEs (Micro, Small, and Medium Enterprises) in Jambi Province. The goal is to make the government's MSME development and support programs more efficient and well-targeted. The data used in this study was obtained from the Department of Cooperatives and MSMEs of Jambi Province, covering the years 2019 to 2023, and includes all 11 regencies and cities in the province. The information collected includes the number of MSMEs (Micro, Small, and Medium), the number of workers, assets, and revenue (omzet) in each region.

To illustrate the development and comparison of MSMEs across regions during the research period, the data are presented in Tables 1 and 2. In the manuscript presentation, only the data from 2019 and 2023 are displayed because these two periods are considered sufficient to comprehensively represent the initial and final conditions of the study. Meanwhile, data from the other years were still utilized in the analysis and data processing stages but were not presented in full in order to maintain the effectiveness and efficiency of the presentation of the research results.

Table 1. MSME Data for the 2019 Period

NO	REGENCY/CITY	BUSINESS			TOTAL	NUMBER OF EMPLOYESS	ASSETS (Rp)	TURNOVER (Rp)
		MICRO	SMALL	MEDIUM				
1	Jambi City	7,257	3,506	-	10,763	21,613	215,260,000,000	233,134,200,000
2	Batanghari Regency	3,849	1,196	45	5,090	12,658	175,431,787,141	432,114,685,593
3	Muaro Jambi Regency	1,297	459	1	1,757	5,924	62,857,251,000	263,129,840,000
4	Tanjab Barat Regency	6,976	1,037	-	8,013	10,437	103,506,819,705	5,640,650,000
5	Tanjab Timur Regency	53,884	1,869	249	56,002	55,825	22,550,931,915	39,333,020,085
6	Tebo Regency	910	268	233	1,411	5,689	83,586,000,000	1,122,489,800,000
7	Bungo Regency	4,291	2,177	380	6,848	16,142	86,366,000,000	43,351,500,000
8	Sarolangun Regency	3,739	564	35	4,338	9,392	153,665,500,000	23,994,721,891
9	Merangin Regency	2,584	677	13	3,274	6,506	270,435,773,000	853,135,215,951
10	Kerinci Regency	30,912	1290	-	32,202	32,202	70,642,715,000	108,265,420,710
11	Sungai Penuh City	7,164	1,127	184	8,772	12,954	709,045,650,441	706,209,865,615
	TOTAL	**123,160**	**14,170**	**1,140**	**138,470**	**188,947**	**1,923,838,427,797**	**3,830,798,829,845**
	YEAR 2018	**90,845**	**12,402**	**605**	**104,155**	**184,124**	**1,693,264,301,121**	**2,347,296,408,135**
	GROWTH (%)	**35.57**	**14.26**	**25.55**	**32.95**	**2.64**	**13.59**	**63.20**

Based on the 2019 data, Jambi Province had a total of 138470 MSMEs, consisting of 123838 micro enterprises, 14170 small enterprises, and 1140 medium enterprises. These MSMEs employed 188947 workers, with total assets of IDR 1.92 trillion and a turnover of IDR 3.83 trillion. Tanjab Timur Regency recorded the highest number of MSMEs, while Sungai Penuh City had the lowest. Compared to the previous year, MSMEs in 2019 increased by 32.95%, with turnover growth reaching 63.20%. In 2020, the number of MSMEs in Jambi Province decreased to 72,126 business units. The number of employed workers also declined to 101972 people. However, MSME assets increased to IDR 2.48 trillion, while turnover reached IDR 4.09 trillion. This condition indicates that economic changes affected the number of business actors, although several MSME sectors were still able to maintain and even increase their assets and turnover. These data are important for identifying changes in MSME patterns during the economic transition period. In 2021, the number of MSMEs increased significantly to 163964 business units, employing 289569 workers. Total MSME assets reached IDR 8.03 trillion, while turnover increased to IDR 309.08 billion. Jambi City recorded the highest number of MSMEs during this period. This substantial increase indicates the recovery of economic activities and MSME growth following the decline in the previous year. Based on the 2022 data, the number of MSMEs in Jambi Province increased to 184042 business units, consisting of 171852 micro enterprises, 9849 small enterprises, and 2341 medium enterprises. These MSMEs employed 307831 workers, with total assets of IDR 3.56 trillion and turnover reaching IDR 39.40 trillion. The data indicate that the MSME sector continued to grow and remained one of the main drivers of the regional economy.

Table 2. MSMES Data for the 2023 Period

NO	REGENCY/CITY	BUSINESS			TOTAL	NUMBER OF EMPLOYESS	ASSETS (Rp)	TURNOVER (Rp)
		MICRO	SMALL	MEDIUM				
1	Merangin Regency	6,840	693	13	7,546	16,103	307,591,958,917	961,405,017,805
2	Jambi City	46,912	3,835	-	50,747	149,629	1,268,675,000,000	11,299,900,000,000
3	Sarolangun Regency	2,510	107	10	2,627	5,292	66,568,925,000	131,350,000,000
4	Sungai Penuh City	7,722	1,125	181	9,028	13,366	528,559,505,438	712,997,686,836
5	Kerinci Regency	6,753	646	69	7,468	12,187	239,492,300,000	258,308,000,000
6	Batang Hari Regency	17,466	138	69	17,673	35,324	191,299,020,000	877,777,279,600
7	Bungo Regency	2,443	881	290	3,614	3,135	32,254,762,500	69,844,393,000
8	Tanjab Barat Regency	7,650	1,048	-	8,698	10,806	93,763,559,200	129,256,657,000
9	Tebo Regency	8,370	-	-	8,370	755	11,401,818,300	2,484,823,380
10	Muaro Jambi Regency	41,234	-	-	41,234	119,674	414,660,000,000	1,484,424,000,000
11	Tanjab Timur Regency	17,658	1,135	253	19,046	7,388	240,706,400,000	23,383,301,440,000
	TOTAL	**165,558**	**9,608**	**885**	**176,051**	**373,659**	**3,394,973,249,355**	**39,311,049,297,621**
	YEAR 2022	**171,852**	**9,849**	**2,341**	**184,042**	**307,831**	**3,560,920,952,281**	**39,405,757,934,297**
	GROWTH (%)	**-4**	**-2**	**-62**	**-4**	**1**	**-5**	**0**

The data used in this study were obtained from MSME statistics in Jambi Province for 2023. The dataset includes information on micro, small, and medium enterprises, total businesses, employees, assets, and turnover across regencies and cities in Jambi Province. In 2023, Jambi Province recorded 176051 MSMEs with 373659 employees, total assets of IDR 3.39 trillion, and turnover of IDR 39.31 trillion. The data also show differences in MSME conditions among regions, particularly in the number of businesses, assets, and turnover. These variables were used as input data in the K-Means clustering process to group regions based on similar MSME characteristics for policy strengthening purposes.

The MSME data from 2019–2023 show dynamic changes in the MSME sector in Jambi Province. In 2020, the number of MSMEs and workers declined due to economic conditions, while significant recovery was observed during 2021–2023 through increases in businesses, employment, assets, and turnover.

The data also indicate differences in MSME characteristics among regencies and cities, particularly in the number of businesses, assets, and turnover. Therefore, these variables were selected as input attributes for the K-Means clustering process to group regions based on similar MSME conditions and support data-driven policy strengthening.

3.2. Preprocessing

The data pre-processing stage was conducted to clean, prepare, and ensure the quality of the data before applying the K-Means clustering algorithm. At this stage, the data were examined to ensure the consistency and completeness of the variables used in the study. The Initial MSME Dataset used in this research is presented in Table 3.

Table 3. MSME Initial Dataset

NO	REGENCY/CITY	BUSINESS			TOTAL	ASSETS (Rp)	TURNOVER (Rp)	TAHUN
		MICRO	SMALL	MEDIUM				
1	Jambi City	7.257	3.506	-	10.763	215.260.000.000	233.134.200.000	2019
2	Batanghari Regency	3.849	1.196	45	5.090	175.431.787.141	432.114.685.593	2019
3	Muaro Jambi Regency	1.297	459	1	1.757	62.857.251.000	263.129.840.000	2019
4	Tanjab Barat Regency	6.976	1.037	-	8.013	103.506.819.705	5.640.650.000	2019
5	Tanjab Timur Regency	53.884	1.869	249	56.002	22.550.931.915	39.333.020.085	2019
….	…..	…..	…..	…..	…..	…..	…..	…..
….	…..	…..	…..	…..	…..	…..	…..	…..
51	Bungo	2.443	881	290	3.614	32.254.762.500	69.844.393.000	2023
52	Tanjung Jabung Barat	7.650	1.048	-	8.698	93.763.559.200	129.256.657.000	2023
53	Tebo	8.370	-	-	8.370	11.401.818.300	2.484.823.380	2023
54	Muaro Jambi	41.234	-	-	41.234	414.660.000.000	1.484.424.000.000	2023
55	Tanjung Jabung Timur	17.658	1.135	253	19.046	240.706.400.000	23.383.301.440.000	2023

Table 3 above presents the MSME dataset in Jambi Province used in this study. The preprocessing stage was carried out as a process of cleaning and preparing raw data before further analysis was conducted. The dataset consists of MSME recapitulation data in Jambi Province from 2019 to 2023, including information on assets, turnover, and annual data for each regency and city.

At this stage, attribute selection was performed to determine the variables relevant to the clustering process. The selected attributes include regency/city name, assets, turnover, and year. Meanwhile, attributes considered irrelevant, such as serial numbers, the number of micro enterprises, small enterprises, medium enterprises, total MSMEs, and employment data, were removed to improve the effectiveness and efficiency of the analysis. The process of removing unnecessary attributes is illustrated in Figure 2.

```
Python

df2 = df.drop(["No", "Micro", "Small", "Medium", "Total", "Employment"], axis=1)
```

Figure 2. The Process Of Removing Unnecessary Attributes

Table 4 above presents the preprocessing results after the unnecessary attributes had been removed. This stage also involved data cleaning processes, including missing value checking, data format standardization, and numerical data conversion to ensure compatibility with the K-Means algorithm. Missing values represented by the symbol "-" were replaced with 0 to ensure that the clustering calculations could be performed properly.

After completing the preprocessing stage, the dataset became cleaner, more structured, and ready for the data transformation and clustering processes using the K-Means algorithm, as presented in Table 4.

Table 4. The Preprocessing Results After The Unnecessary Attributes Had Been Removed

NO	REGENCY/CITY	TOTAL	ASSETS (Rp)	TURNOVER (Rp)	TAHUN
1	Jambi City	10.763	215.260.000.000	233.134.200.000	2019
2	Batanghari Regency	5.090	175.431.787.141	432.114.685.593	2019
3	Muaro Jambi Regency	1.757	62.857.251.000	263.129.840.000	2019
4	Tanjab Barat Regency	8.013	103.506.819.705	5.640.650.000	2019
5	Tanjab Timur Regency	56.002	22.550.931.915	39.333.020.085	2019
….	…..	…..	…..	…..	…..
….	…..	…..	…..	…..	…..
51	Bungo	3.614	32.254.762.500	69.844.393.000	2023
52	Tanjung Jabung Barat	8.698	93.763.559.200	129.256.657.000	2023
53	Tebo	8.370	11.401.818.300	2.484.823.380	2023
54	Muaro Jambi	41.234	414.660.000.000	1.484.424.000.000	2023
55	Tanjung Jabung Timur	19.046	240.706.400.000	23.383.301.440.000	2023

Table 4 presents the preprocessing results after the unnecessary attributes had been removed from the dataset. The table contains relevant attributes used in the clustering process, namely regency/city, total MSMEs, assets, turnover, and year. The data represent MSME conditions in each regency/city in Jambi Province from 2019 to 2023. The preprocessing stage produced a cleaner and more structured dataset, making it suitable for further analysis using the K-Means clustering algorithm.

3.3. Transformation

The data transformation stage was carried out to modify the data structure in order to facilitate data processing according to the requirements of the data mining process. At this stage, the data, which were previously stored in a single column representing the number of MSMEs for each year, were transformed into multiple columns grouped by year. The transformation process separated the MSME data into the 2019, 2020, 2021, 2022, and 2023 columns. In addition, a total attribute was added to represent the overall number of MSMEs from 2019 to 2023.

The transformation process produced 11 regency/city records in Jambi Province, which are listed in the Regency/City column. Each regency/city contains MSME data for each year from 2019 to 2023, presented in the corresponding yearly columns. The final column, namely total, represents the cumulative number of MSMEs over the five-year period. Table 5 presents the results of the MSME data transformation process in Jambi Province.

Table 5. MSME Data in Jambi Province

NO	REGENCY/CITY	2019	2020	2021	2022	2023	TOTAL
1	Jambi City	10763	10763	47813	50747	50747	170833
2	Sungai Penuh City	8772	8772	8098	9631	9028	44301
3	Batanghari Regency	5090	4062	12796	17611	17673	57232
4	Muaro Jambi Regency	1757	1757	42105	41234	41234	128087
5	Tanjab Barat Regency	8013	8110	8390	8698	8698	41909
6	Tanjab Timur Regency	56002	13220	19046	19046	19046	126360
7	Tebo Regency	1411	2093	1268	8370	8370	21512
8	Bungo Regency	6848	3387	3387	12489	3614	29725
9	Sarolangun Regency	4338	4338	3705	2627	2627	17635
10	Merangin Regency	3274	3554	4956	6121	7546	25451
11	Kerinci Regency	32202	12070	12400	7468	7468	71608

Table 5 presents MSME data in Jambi Province by regency/city during the 2019–2023 period. The table illustrates the distribution and development of MSMEs in each region over five years, including the total number accumulated during the observation period. Based on the table, Jambi City recorded the highest number of MSMEs with a total of 170,833 units, followed by Muaro Jambi Regency with 128,087 units and Tanjab Timur Regency with 126,360 units. In contrast, Tebo Regency had the lowest number of MSMEs, totaling 21,512 units. Overall, the data indicate fluctuations in MSME numbers across several regencies/cities, although most regions showed an increasing trend in the later years. The longitudinal analysis revealed that the COVID-19 pandemic in 2020 significantly affected MSME stability across several regencies/cities in Jambi Province. However, most regions gradually recovered during 2021–2023, indicating that the cluster structure remained relatively stable despite temporary economic disruptions.

3.4. K-Means Clustering Process

This stage involves processing the data to generate meaningful information and knowledge using a specific analytical method. In this study, the clustering process was performed using the K-Means algorithm, implemented through the Python programming language.

This study employs the K-Means algorithm as an unsupervised machine learning approach to identify cluster structures within MSME data in Jambi Province. As a clustering method without predefined class labels, K-Means groups objects based on the similarity of their characteristics, enabling the identification of natural patterns that may not be visible through conventional descriptive analysis. Therefore, this approach places the study within the scope of Artificial Intelligence and Machine Learning (AI/ML), as the analysis process not only describes the data but also extracts hidden information through data-driven learning.

The use of K-Means in this study is considered relevant because MSME data exhibit diverse characteristics, particularly in terms of assets, turnover, and business scale variables. Through the clustering process, administrative data that were previously scattered can be transformed into more homogeneous and interpretable groups. The resulting clusters can then be used as a basis for understanding differences in MSME conditions across regions and for supporting more targeted policy formulation.

Prior to the clustering process, the dataset was normalized using the Min-Max Scaling technique to reduce differences in variable ranges and improve clustering performance. In addition, outlier detection was conducted using the Interquartile Range (IQR) method to identify extreme values that could potentially distort cluster boundaries in the K-Means algorithm.

The initial step in the K-Means clustering process is determining the optimal number of clusters to be formed. To achieve this, the Elbow Method was employed as an optimization approach for selecting the most appropriate number of clusters. The purpose of the Elbow Method is to identify the smallest value of k that produces a low within-cluster variation(Alamanda et al., 2023).

The optimal cluster value is determined by comparing the SSE (Sum of Squared Errors) values for different numbers of clusters. The appropriate number of clusters is identified at the point where the graph forms an "elbow," indicating that additional clusters no longer provide significant improvement in reducing SSE values. Figure 3 illustrates the Elbow Method graph used to determine the optimal number of clusters in this study.

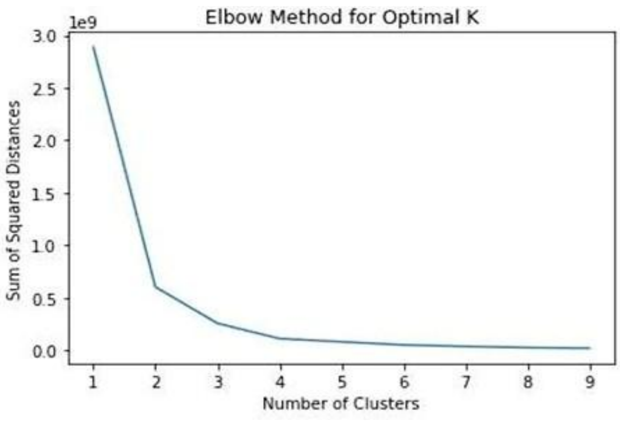

Figure 3. Elbow Method Result Graph

Figure 3 illustrates the results of the Elbow Method used to determine the optimal number of clusters in the dataset. The graph shows a significant decrease in the SSE (Sum of Squared Errors) value from k=1 to k=3, after which the decline becomes more gradual. The "elbow" point appears at k=3, indicating that three clusters provide the optimal grouping for the data. Therefore, the study adopted three clusters for the K-Means clustering process.The determination of k=3 was further validated using the Silhouette Coefficient and Davies-Bouldin Index (DBI), following the recommendation of Yuan and Yang (2019), which emphasized combining multiple validation approaches to improve clustering reliability for high-dimensional economic datasets.

A total of three clusters were formed in this study, namely the low cluster (C0), which represents regions with a low distribution of MSMEs, the medium cluster (C1), which represents regions with a moderate distribution of MSMEs, and the high cluster (C2), which represents regions with a high distribution of MSMEs(Vásquez et al., 2021). Regions categorized within the Low Cluster (C0) generally face socio-economic challenges such as limited digital infrastructure, lower access to financing, restricted market accessibility, and lower entrepreneurial capacity development. These conditions potentially hinder MSME competitiveness and regional economic growth.The initial step in the K-Means clustering process involves determining the initial centroids. These centroids were established based on the minimum value for C0, the average value for C1, and the maximum value for C2(Harahap et al., 2022). The clustering process was conducted using only the total number attribute as the primary variable for grouping the data. Table 6 presents the initial centroid values used in the dataset.

Table 6. Initial Centroid Values

Centroid	
Min (C0)	10972
Average (C1)	78571,48
Max (C2)	154334

Table 6 presents the initial centroid values used in the K-Means clustering process. The centroids were determined based on three representative values: the minimum value for Cluster 0 (C0), the average value for Cluster 1 (C1), and the maximum value for Cluster 2 (C2). The minimum centroid value of 10972 represents the low MSME distribution cluster, the average centroid value of 78,571.48 represents the medium MSME distribution cluster, and the maximum centroid value of 154334 represents the high MSME distribution cluster. These centroid values served as the initial reference points for grouping the data into clusters.

Tabel 7. Final Centroid Values

Final Centroid	
C0 (Merangin, Sarolangun, Tebo)	21,532.67
C1 (Bungo, Tanjab Barat, Tanjab Timur, Sungai Penuh)	60,573.75
C2 (Batanghari, Kerinci, Muaro Jambi, Jambi City)	106,940.00

Table 7 presents the final centroid values obtained after the K-Means clustering iteration process was completed. The final centroid values represent the average characteristics of MSME distribution within each cluster. Cluster C0 has the lowest centroid value, indicating regions with relatively lower MSME distribution levels. Meanwhile, Cluster C2 has the highest centroid value, representing regions with more developed MSME distribution characteristics. These centroid values indicate the differences in MSME distribution intensity among regencies and cities in Jambi Province.

After determining the initial centroids, the next step was to perform the clustering process by calculating the closest distance between each data point and the centroids(Sari et al., 2026). The clustering results obtained using the Python tool can be seen in Figure 4.

Based on Figure 4, the clustering results show the grouping of regencies and cities in Jambi Province into three cluster categories based on the characteristics of MSME distribution. The Low Cluster (C0) consists of Merangin Regency, Sarolangun Regency, and Tebo Regency, indicating that these regions have relatively lower levels of MSME distribution compared to other regions.

The Medium Cluster (C1) includes Bungo Regency, Tanjab Barat Regency, Tanjab Timur Regency, and Sungai Penuh City. These regions demonstrate a moderate level of MSME distribution and represent areas with relatively balanced MSME development.

Regency/City	Cluster	Cluster_Category
Batanghari Regency	C2	High Cluster
Bungo Regency	C1	Medium Cluster
Kerinci Regency	C2	High Cluster
Merangin Regency	C0	Low Cluster
Muaro Jambi Regency	C2	High Cluster
Sarolangun Regency	C0	Low Cluster
Tanjab Barat Regency	C1	Medium Cluster
Tanjab Timur Regency	C1	Medium Cluster
Tebo Regency	C0	Low Cluster
Jambi City	C2	High Cluster
Sungai Penuh City	C1	Medium Cluster

Figure 4. Elbow Method Result Graph

Meanwhile, the High Cluster (C2) consists of Batanghari Regency, Kerinci Regency, Muaro Jambi Regency, and Jambi City, indicating that these regions have higher levels of MSME distribution compared to the other clusters. Overall, the clustering results reveal differences in MSME distribution patterns among regencies and cities in Jambi Province, which can serve as a reference for regional development strategies and more targeted policy formulation. A cluster-based strategic roadmap is necessary to bridge the gap between clustering analysis and policy implementation. High Cluster (C2) regions should focus on innovation ecosystems, export-oriented MSME strengthening, and digital transformation programs. Medium Cluster (C1) regions require business scaling and market expansion strategies, while Low Cluster (C0) regions need infrastructure support, entrepreneurship mentoring, financial inclusion, and micro-financing assistance.

3.5. valuation

The results of the K-Means modeling process need to undergo an evaluation stage to determine the quality of the clustering results obtained. This evaluation aims to assess whether the formed clusters are optimal and capable of representing the characteristics of the data effectively. In this study, cluster quality evaluation was conducted using two testing methods, namely the Silhouette Coefficient and the Davies-Bouldin Index (DBI).

The Silhouette Coefficient is used to measure the similarity of a data point to its own cluster compared to other clusters. A higher Silhouette Coefficient value indicates better clustering quality. Meanwhile, the Davies-Bouldin Index is used to evaluate cluster quality based on the proximity between clusters and the distribution of data within each cluster. A lower DBI value indicates a more optimal clustering result.

The subjective criteria for measuring clustering quality based on the Silhouette Coefficient method according to Kaufman and Rousseeuw (1990) are presented in Figure 5.

Table 8. Silhouette Coefficient Evaluation Criteria

SC Value	Criteria
0.71 – 1.00	Strong Structure
0.51 – 0.70	Good Structure
0.26 – 0.50	Weak Structure
≤ 0.25	Poor Structure

Table 8 presents the evaluation criteria for clustering quality based on the Silhouette Coefficient (SC) value. The Silhouette Coefficient is used to measure the degree of similarity of data points within the same cluster compared to other clusters. SC values range from -1 to 1, where higher values indicate better clustering quality.

Based on the criteria shown in the table, an SC value between 0.71 and 1.00 indicates a strong cluster structure, meaning that the clustering results are highly accurate and well separated. An SC value between 0.51 and 0.70 represents a good cluster structure, while values between 0.26 and 0.50 indicate a weak cluster structure. Meanwhile, an SC value of 0.25 or lower reflects a poor cluster structure, suggesting that the clustering results are not well defined. The results of the Silhouette Coefficient testing using Python in this study are presented as follows.

```python
from sklearn.metrics import silhouette_score

# Menghitung nilai Silhouette Coefficient
score = silhouette_score(df2, km.labels_)

# Menampilkan hasil
print("Silhouette Coefficient Score:", score)
```

Figure 5. Silhouette Coefficient Evaluation Results

Based on the testing results shown in Figure 5, the Silhouette Coefficient value obtained was 0.81. This value indicates that the resulting clustering quality falls into the strong structure category, which means that the clusters formed are well defined, highly cohesive, and have a high level of clustering accuracy.

In addition to the Silhouette Coefficient evaluation, the clustering results were also assessed using the Davies-Bouldin Index (DBI) method to further measure the quality and separation of the clusters formed.

```python
from sklearn.metrics import davies_bouldin_score

# Menghitung nilai Davies-Bouldin Index (DBI)
dbi_score = davies_bouldin_score(df2, km.labels_)

# Menampilkan hasil
print("Davies-Bouldin Index (DBI):", dbi_score)
```

Figure 6. Davies-Bouldin Index (DBI) Evaluation Results

Based on the testing results shown in Figure 6, the Davies-Bouldin Index (DBI) value obtained was 0.21. This value indicates that the resulting clustering model has good quality, since in DBI evaluation, a value closer to zero reflects a more optimal clustering performance.

A DBI value of 0.21 indicates that the overlap between clusters is relatively low and the distance between data groups is sufficiently clear. This result shows that each cluster has distinct characteristics and is capable of effectively representing the distribution patterns of MSMEs in Jambi Province. Therefore, the evaluation results using the DBI method indicate that the K-Means clustering implemented in this study falls into the good category.

Based on the comparison of the two evaluation methods, namely the Silhouette Coefficient and the Davies-Bouldin Index, both methods demonstrated good evaluation results for the implementation of K-Means clustering in the case of MSME distribution in Jambi Province. These findings indicate that the clustering model used in this study was able to produce optimal and representative data groupings.

4. Conclusion

This study successfully applied the K-Means algorithm as an unsupervised machine learning method to analyze the distribution of MSMEs in Jambi Province based on data from 2019–2023. The K-Means method was used to group data based on the similarity of characteristics, allowing the identification of MSME distribution patterns that may not be visible through conventional descriptive analysis. Therefore, this study is included within the application of Artificial Intelligence and Machine Learning (AI/ML) because it is able to extract information from data automatically through a data-driven learning process. The use of the K-Means algorithm is considered appropriate because MSME data have diverse characteristics, particularly in terms of assets, turnover, and business scale variables. Through the clustering process, scattered administrative data can be grouped into more homogeneous and interpretable categories. The clustering results can then be used to understand differences in MSME conditions across regions and support more targeted policy-making. The clustering process was carried out using the Python programming language and optimized using the Elbow Method, which produced three optimal clusters: Low Cluster (C0), Medium Cluster (C1), and High Cluster (C2). The clustering results showed that Merangin Regency, Sarolangun Regency, and Tebo Regency were included in the Low Cluster (C0). Bungo Regency, Tanjung Jabung Barat Regency, Tanjung Jabung Timur Regency, and Sungai Penuh City were included in the Medium Cluster (C1). Meanwhile, Batanghari Regency, Kerinci Regency, Muaro Jambi Regency, and Jambi City were included in the High Cluster (C2). The evaluation results using the Silhouette Coefficient and Davies-Bouldin Index (DBI) showed that the clustering model produced good quality clusters. The Silhouette Coefficient value of 0.81 indicated a strong cluster structure, while the DBI value of 0.21 indicated that the resulting clusters were optimal with low overlap between clusters. Therefore, the K-Means model implemented in this study is considered effective in identifying MSME distribution patterns and can support regional development planning and MSME policy formulation in Jambi Province. Furthermore, cluster-specific policy implementation can contribute to reducing the urban-rural economic divide in Jambi Province by enabling more adaptive and region-oriented MSME development strategies based on local socio-economic characteristics.

Acknowledgement

The authors would like to express their sincere gratitude to all parties who contributed to the completion of this research. Special appreciation is extended to the Regional Office of Cooperatives and Small Enterprises of Jambi for providing the MSME data and supporting information used in this study. The authors also appreciate the support, guidance, and valuable suggestions provided by colleagues and academic supervisors throughout the research process. These contributions greatly supported the successful completion of this study.

References

Alamanda, D. T., Kusmiati, E., & Roji, F. F. (2023). MSME Clusterization Using K-Means Clustering in Garut Regency, Indonesia. *Review of Integrative Business and Economics Research*, *12*(4), 199.

Atemoagbo, O. P., Abdullahi, A., & Siyan, P. (2024). Cluster Analysis of MSMES In Suleja, Nigeria: Insights From Fuzzy C-Means Clustering And T-SNE Visualizations. *Management and Economic Journal*, *2024*(1), 1–9. https://doi.org/10.18535/mej/v2023.03

Bharadiya, J. (2023). The Impact of Artificial Intelligence on Business Processes. *European Journal of Technology*, *7*(2), 15–25. https://doi.org/10.47672/ejt.1488

Boubaker, S., Le, T. D. Q., Ngo, T., & Manita, R. (2025). Predicting the performance of MSMEs: a hybrid DEA-machine learning approach. *Annals of Operations Research*, *350*(2), 555–577. https://doi.org/10.1007/s10479-023-05230-8

Cao, W. (2025). Research on Credit Risk Quantification and Optimization Strategies for Micro, Small and Medium Enterprises (MSMEs). In *Proceedings of 2025 4th International Conference on Cyber Security, Artificial Intelligence and the Digital Economy, CSAIDE 2025* (Vol. 1, Issue 1). Association for Computing Machinery. https://doi.org/10.1145/3729706.3729727

Caraka, R. E., Kurniawan, R., Nasution, B. I., Jamilatuzzahro, J., Gio, P. U., Basyuni, M., & Pardamean, B. (2021). Micro, small, and medium enterprises' business vulnerability cluster in indonesia: An analysis using optimized fuzzy geodemographic clustering. *Sustainability (Switzerland)*, *13*(14). https://doi.org/10.3390/su13147807

Ezeudo, M.T. and Biose, A. O. (2024). Impacts of Small and Medium Enterprises on Economic Growth in. *COEASU Erudite Journal*, *6*(1), 365–380.

Harahap, M., Zamili, A. W. D. R., Arvansyah, M. A., Saragih, E. F., Rajen, S., & Husein, A. M. (2022). K-Means Clustering Algorithm Approach in Clustering Data on Cocoa Production Results in the Sumatra Region. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *6*(6), 905–910. https://doi.org/10.29207/RESTI.V6I6.4199

Kim, J. S., & Seo, D. (2023). Foresight and strategic decision-making framework from artificial intelligence technology development to utilization activities in small-and-medium-sized enterprises. *Foresight*, *25*(6), 769–787. https://doi.org/10.1108/FS-06-2022-0069

Kondoj, M. A. S., Waroh, A., Walukow, S., & Pairunan, T. (2025). *Improving the Effectiveness of MSME Marketing through Data Mining Integration In CRM Systems*. *2025*, 778–786. https://doi.org/10.2991/978-94-6463-926-

1_87

Kustiyahningsih, Y., Khozaimi, A., Khotimah, B. K., Ainiyah, A., Sari, M. M., Maghfiroh, I., Insani, A. N., & Lutfiyah, R. D. (2024). Integration K-Means clustering and AHP for recommendations batik MSMEs. *E3S Web of Conferences*, *499*, 1–7. https://doi.org/10.1051/e3sconf/202449901006

Lestari, R., Lestari, F., Saptiyanto, N. D., & daffansyah, C. (2025). Technology Readiness Profiling of SMEs With K-Means Clustering in Bandung, Indonesia. *Jurnal Ekonomi Bisnis Dan Manajemen (Eko-Bisma)*, *4*(2), 418–425. https://doi.org/10.58268/eb.v4i2.221

Maksum, I. R., Sri Rahayu, A. Y., & Kusumawardhani, D. (2020). A social enterprise approach to empowering micro, small and medium enterprises (SMEs) in Indonesia. *Journal of Open Innovation: Technology, Market, and Complexity*, *6*(3). https://doi.org/10.3390/JOITMC6030050

Maulidya, A. M., Selfira, S., Sidabutar, G., & Al Hafiz, R. R. (2025). Integrating K-Means Clustering and Apriori for Data Mining-Based Digital Marketing Strategy For Increasing UMKM: Study Case Stabat City. *Sinkron*, *9*(4), 3190–3196. https://doi.org/10.33395/sinkron.v9i4.15299

Murdiono, A., Pratikto, H., Handayati, P., Winarno, A., Subagyo, & Zen, F. (2026). Analysis of MSME Financial Literacy Segmentation Using Clustering Methods. *International Journal on Informatics Visualization*, *10*(2), 900–908. https://doi.org/10.62527/joiv.10.2.4743

Nouvel, A., & Dwi Mulyanto, J. (2025). Clustering of MSMEs Based on Assets and Turnover Using the K-Means Algorithm-Supriatiningsih et.al Clustering of MSMEs Based on Assets and Turnover Using the K-Means Algorithm. *Jurnal Multidisiplin Sahombu*, *5*(07), 2204–2212. https://doi.org/10.58471/jms.v5i07

Nugraha, R. S., Jatiningrum, W. S., & Astuti, R. D. (2021). Cluster analysis to determine business strategy for MSMEs in Yogyakarta. *IOP Conference Series: Materials Science and Engineering*, *1072*(1), 012011. https://doi.org/10.1088/1757-899x/1072/1/012011

Pamungkas, M. R. S. P., Asyivadibrata, A., Susilawati, T., & Huda, M. N. (2023). Unleashing the Potentials of Artificial Intelligence for Micro, Small, and Medium Enterprises: A Systematic Literature Review. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *5*(3), 303–310. https://doi.org/10.47233/jteksis.v5i3.860

Ragazou, K., Passas, I., Garefalakis, A., & Zopounidis, C. (2023). Business intelligence model empowering SMEs to make better decisions and enhance their competitive advantage. *Discover Analytics*, *1*(1). https://doi.org/10.1007/s44257-022-00002-3

Ramadhan, A., Achmad, F., Zulkarnain, I., & Aritsugi, M. (2025). Evaluation of K-Means, DBSCAN, and Hierarchical Clustering for Strategic Segmentation of Tourism SMEs in Rembang, Indonesia. *Jurnal Teknik Informatika (Jutif)*, *6*(3), 1605–1630. https://doi.org/10.52436/1.jutif.2025.6.3.4602

Reyes-Ruiz, G., & Hernandez-Hernandez, M. (2021). Fuzzy clustering as a new grouping technique to define the business size of SMEs through their financial information. *Journal of Intelligent and Fuzzy Systems*, *40*(2), 1773–1782. https://doi.org/10.3233/JIFS-189184

Sari, L., Maharrani, R. H., Hastuti, H. D., & Ramadhan, A. P. (2026). *Mixed-Data K-Means Clustering with Hyperparameter-Tuned Random Forest for OSS-Based MSME Investment Profiling and Policy Targeting*. *7*(2), 1535–1548.

Stevenson, M., Mues, C., & Bravo, C. (2021). The value of text for small business default prediction: A Deep Learning approach. *European Journal of Operational Research*, *295*(2), 758–771. https://www.sciencedirect.com/science/article/abs/pii/S0377221721001983

Sutramiani, N. P., Arthana, I. M. T., Lampung, P. F., Aurelia, S., Fauzi, M., & Darma, I. W. A. S. (2024). The Performance Comparison of DBSCAN and K-Means Clustering for MSMEs Grouping based on Asset Value and Turnover. *Journal of Information Systems Engineering and Business Intelligence*, *10*(1), 13–24. https://doi.org/10.20473/jisebi.10.1.13-24

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, *336*(1). https://doi.org/10.1088/1757-899X/336/1/012017

Terttiaavini, T. (2024). A Hybrid Approach Using K-Means Clustering and the SAW Method for Evaluating and

Determining the Priority of SMEs in Palembang City. *Journal of Intelligent System and Computation*, *6*(1), 46–53. https://doi.org/10.52985/insyst.v6i1.392

Vásquez, J., Aguirre, S., Puertas, E., Bruno, G., Priarone, P. C., & Settineri, L. (2021). A sustainability maturity model for micro, small and medium-sized enterprises (MSMEs) based on a data analytics evaluation approach. *Journal of Cleaner Production*, *311*(May). https://doi.org/10.1016/j.jclepro.2021.127692

Viloria, A., & Lezama, O. B. P. (2019). Improvements for determining the number of clusters in k-means for innovation databases in SMEs. *Procedia Computer Science*, *151*(2018), 1201–1206. https://doi.org/10.1016/j.procs.2019.04.172

Wibowo, N. L., Soeleman, M. A., & Fanani, A. Z. (2023). Antlion Optimizer Algorithm Modification for Initial Centroid Determination in K-means Algorithm. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *7*(4), 870–883. https://doi.org/10.29207/RESTI.V7I4.4997

Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. *J*, *2*(2), 226–235. https://doi.org/10.3390/j2020016