JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 1-13
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.760

MSME Segmentation in Pekanbaru Based on Local E-Catalog Participation Using K-Means

Rahma Aliya[1*], Inggih Permana[2], Febi Nur Salisah[3], Rice Novita[4], Muhammad Jazman[5]

[1,2,3,4,5]Faculty of Science and Technology, Information Systems, University Sultan Syarif Khasim, Riau, Indonesia
Email: [1*]12250320351@students.uin-suska.ac.id, [2]inggihpermana@uin-suska.ac.id, [3]febinursalisah@uin-suska.ac.id, [4]rice.novita@uin-suska.ac.id, [5]jazman@uin-suska.ac.id
(* Email Corresponding Author: 12250320351@students.uin-suska.ac.id)
Received: January 3, 2026. | Revision: February 1, 2026 | Accepted: March 3, 2026

Abstrahact

Micro, Small, and Medium Enterprises (MSMEs) play a vital role in the economy; however, their participation in digital government procurement platforms such as the Local E-Catalog in Pekanbaru City remains relatively low. The lack of comprehensive, data-driven mapping of MSME characteristics has resulted in less targeted development and assistance programs. This study aims to segment MSMEs based on revenue, number of employees, and participation status in the Local E-Catalog to generate business groups that can support more effective development strategies. A data mining approach using the K-Means clustering algorithm was applied and implemented through the Orange Data Mining application. The results indicate that a three-cluster configuration is the most optimal, achieving the highest Silhouette Score of 0.444. Cluster 1 represents micro-scale MSMEs with low business capacity and minimal participation in the Local E-Catalog, Cluster 2 consists of growing MSMEs with moderate business capacity, and Cluster 3 comprises established MSMEs with high business capacity and active participation in the Local E-Catalog. These findings provide empirical evidence to support local governments in formulating more targeted and data-driven policies for accelerating MSME digitalization.

Keywords: *Clustering, Local E-Catalouge, K-Means, Segmentasi, MSMEs*

1. INTRODUCTION

Micro, Small, and Medium Enterprises (MSMEs) are individual business entities that play a strategic role in the national economy. Although small in scale, MSMEs are capable of producing competitive products that are in demand in international markets, thereby indirectly contributing to increasing export value and strengthening Indonesia's trade balance. [1]. In addition, MSMEs also constitute one of the main pillars of Indonesia's economic structure, as their presence is widespread across all regions, reflecting economic activity at the grassroots level. Through their contribution to absorbing labor, increasing community income, and promoting equitable development, MSMEs play an important role in maintaining stability and driving overall national economic growth [2].

Micro, small, and medium enterprises (MSMEs) play an important role in absorbing more than 97% of the workforce in Indonesia, which accounts for 57% of the gross domestic product (GDP) . The existence of MSMEs in Indonesia's economic structure therefore occupies a highly significant position [3]. In the context of MSME development, many actors have begun to adopt the use of information technology, which can increase the efficiency and effectiveness of existing processes. On a broader scale, the use of digital technology can improve overall business performance and strengthen the ability of MSMEs to obtain and deliver information to the market

One form of the Indonesian government's efforts to utilize technology as a medium for MSME development is the use of electronic catalogs in the procurement of public goods and services. Electronic catalogs, as part of electronic commerce processes, are believed to be capable of opening up new markets. An electronic catalog is one component of an e-marketplace in the procurement of goods and services, namely an electronic marketplace provided to meet the needs of government goods and services LKPP, 2018. Given the diversity of market coverage and characteristics, electronic catalogs are classified into several types: national, sectoral, and local electronic catalogs. The national electronic catalog is compiled and managed by the National Public Procurement Agency, the sectoral electronic catalog contains information on general and innovative goods/services managed by ministries or agencies, and the regional or local electronic catalog is managed by local governments

Local or regional e-catalogs have emerged as one of the strategic solutions to minimize fraudulent practices in the procurement of goods and services. However, despite offering transparency and efficiency, the adoption rate of e-catalogs among business actors is still relatively low. In fact, this platform offers various benefits, especially for policymakers and those with authority in the procurement process, because it can speed up procedures, increase accountability, and reduce the potential for irregularities[4].

However, MSME participation in government digital platforms such as e-catalogs remains low, even though this system is crucial for improving transparency and efficiency in the procurement of public goods and services; the limited participation creates a digital divide and restricts market opportunities for local MSMEs, and although strategic, MSME

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 1-13
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.760

involvement in government digital systems such as local e-catalogs is still very limited because many MSMEs have not been integrated due to constraints in digital literacy and resource capacity[5]. In Pekanbaru City, this condition creates a digital divide that limits the marketing potential of MSMEs through government procurement channels. [6].

Therefore, segmenting MSMEs based on business characteristics is a strategic approach to supporting more accurate and targeted policymaking; to date, government interventions for MSMEs, such as training programs or financial assistance, have tended to be implemented in a general manner without considering digital readiness, business capacity, or the level of participation in e-catalog platforms, leading to mistargeted policies, low program effectiveness, and resource waste, whereas through data-driven segmentation the government can obtain a clearer picture of the profiles, needs, and potential of each business group so that every cluster can receive interventions that align with its actual conditions, making data-based segmentation a crucial foundation for realizing policies that are responsive to the diversity and dynamics of the MSME sector [7].

To understand the characteristics of MSMEs that have not yet participated in local e-catalogs, a data-driven analytical approach is required. Clustering is an analytical approach that is commonly used for data segmentation. One of the most widely used algorithms is K-Means; however, other approaches such as Hierarchical Clustering and DBSCAN are also relevant for providing a comparative perspective in identifying MSME patterns for segmentation based on parameters such as revenue, number of employees, type of product, and digitalization status. [8]. With this approach, the government can formulate appropriate interventions for each MSME group according to their digital readiness and business capacity[9].

The application of K-Means by the K-Means Clustering method is one of the algorithms widely used in data segmentation due to its simple and efficient process, especially in handling large amounts of data. In the context of MSME clustering, this algorithm is applied to classify business units based on criteria such as turnover, number of employees, and type of business. The segmentation results obtained are able to show differences in characteristics between groups, for example, the existence of MSME clusters with high turnover but few employees, and vice versa. This information can be used to design more targeted coaching and policy development strategies, in accordance with the needs of each business segment) [10].

Previous studies have shown the effectiveness of the K-Means Clustering algorithm in segmenting MSMEs and various other business entities. The study grouped MSME store products based on restocking needs with a Davies-Bouldin Index (DBI) value of 0.436, indicating fairly good segmentation. Research by Mawarni et al. (2023) also shows that the K-Means method is capable of grouping customers based on purchase value, resulting in priority customer groups that are useful for marketing strategies. Similar findings were reported by [11] which applied K-Means for MSME market segmentation with a Silhouette Score of 0.72 and a DBI of 0.45, indicating a stable and representative model. Meanwhile, Arrahmi and Terttiaavini applied K-Means in the context of regional MSMEs using variables such as industry sector, number of workers, and turnover. Their research successfully formed clusters of micro, small, and medium enterprises with different economic characteristics, providing a basis for the formulation of more targeted MSME development policies.

On the other hand, a number of studies highlight the importance of digitization through government e-catalogues as a means of expanding the MSME market[12] studied the implementation of e-catalogues in Pekanbaru City and found that this system improves the transparency and efficiency of the procurement process for goods and services, but the level of participation by MSMEs remains low due to limited capital, digital literacy, and access to information. Research by [13] Reviewing the regulations and procedures for becoming an e-catalog provider under the LKPP shows that administrative complexity remains a major obstacle. International research by[14] Introducing the Differentiation-based Initialization method to improve the weaknesses of centroid initialization in K-Means, which demonstrates a global effort to improve the accuracy and stability of this method.

Based on a review of previous studies, it can be concluded that studies on the digitization of MSMEs and the application of clustering have contributed significantly to the development of small business strategies. However, most studies are still conducted separately: studies on e-catalogues tend to be descriptive and focus on policy, while studies on clustering generally highlight business characteristics without considering the aspect of digitization.

This is where the research gap that this study aims to fill arises, namely the lack of research that directly links the level of MSME participation in local e-catalogues with the results of K-Means algorithm-based segmentation. This study seeks to fill this gap by analyzing MSME data in Pekanbaru City using a data mining approach to group MSMEs based on turnover, number of employees, length of establishment, and level of participation in local e-catalogues. The evaluation process was carried out using the Silhouette Score to ensure the quality of the clusters formed.

With this approach, this study is expected to produce an accurate and applicable segmentation map of Pekanbaru's digital MSMEs, as well as provide strategic recommendations for local governments in improving digital literacy and MSME participation in e-catalog platforms. The results of this study are also expected to contribute significantly to strengthening the literature on e-procurement integration and data-driven clustering in the context of the digital economy in Indonesia.

This study contributes to the development of evidence-based strategies by mapping MSMEs in Pekanbaru based on their involvement in e-catalog platforms. This dimension has not been widely explored at the local level in Indonesia,

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 1-13
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.760

especially with a data mining-based approach. [15]. This segmentation is expected to form the basis for the formulation of MSME development policies based on digital characteristics, thereby supporting inclusive and sustainable economic transformation.

2. RESEARCH METHODOLOGY

2.1 Type of Research

This study applies a descriptive quantitative approach that emphasizes objectivity and numerical data measurement [16]. A descriptive approach was used to provide a systematic and factual description of the participation of MSMEs in the Local E-Catalog without seeking causal relationships. Specifically, this study falls under exploratory data mining, which utilizes the K-Means Clustering algorithm to find segmentation patterns based on similarities in variable characteristics, thereby producing information that is useful for decision-making strategies.

2.2 Research Process

The research flow describes the systematic stages carried out in this study, from planning to analysis of results. The following is the research flow:

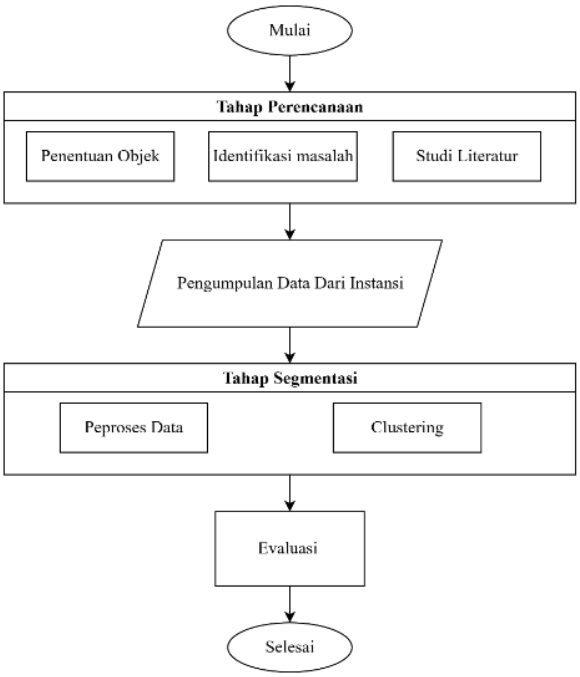

Figure 1. Research Flow

This research was conducted systematically through four main stages: planning, data collection, segmentation, and evaluation.

1. **Planning Stage**

Covering the determination of research objects (MSMEs registered with the Pekanbaru City Cooperative Office), identification of problems related to low participation in the E-Catalog, and a literature study on the digitization of government services and clustering algorithms[12]. The planning stage includes determining the research object, identifying problems, and conducting literature studies as the basis for preparing the direction of the segmentation analysis of MSMEs in Pekanbaru City.

a. Determination of Research Object

The research focuses on 10,560 verified MSMEs in Pekanbaru City based on 2024 data. The objects were selected to support the need to increase MSME participation in the Local e-Catalog. The variables analyzed include turnover, workforce, length of business establishment, and MSME participation status. Segmentation was carried out to identify business groups that are ready, have potential, or require intervention in order to participate in the digital procurement system.

b. Problem Identification

The initial analysis revealed several key issues: low MSME participation in the e-Catalog, inconsistencies in data format and completeness between agencies, heterogeneity of business capacity that has not been systematically mapped, and the lack of a data-based classification mechanism to support MSME development strategies. These

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 1-13
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.760

conditions underscore the need for a clustering approach to produce objective segmentation that can be used as a basis for policy formulation.

c. Literature Review

The literature reviewed includes MSME regulations (Law No. 20/2008), the role of the e-Catalog as a government procurement instrument, the application of the K-Means algorithm in MSME clustering, and the concept of data-based segmentation in regional economic policy. This study strengthens the methodological framework of the research and underpins the selection of variables, preprocessing techniques, and clustering methods used.

2. **Data Collection Stage**

Data collection in this study used secondary data obtained from two government agencies, namely the Pekanbaru City Cooperative, Small and Medium Enterprises (SME) Office and the Electronic Procurement Service (LPSE). Data from the Pekanbaru City Cooperative and SME Office consisted of profiles of 10,560 verified Micro, Small, and Medium Enterprises (MSMEs) in 2024, obtained in spreadsheet format (.xlsx/.csv). Meanwhile, supporting data was obtained from the LPSE of Pekanbaru City, which contains information on the e-Catalog and goods/service providers. The integration of these two data sources enables a comprehensive analysis of the characteristics of MSMEs and their level of participation in the digital ecosystem of government procurement through the e-Catalog platform. The main variables used in this study are:

Table 1. Research variabels

Variable	Variable Type	Used in Clustering
Business Category	Categorical	Yes
Number of Employees	Numerical	Yes
Revenue / Turnover	Numerical	Yes
Business Age	Numerical	Yes
Local E-Catalog Participation	Categorical (Binary)	Yes
MSME Name	Identifier	No

Of these variables, only numerical variables (number of employees, turnover, length of establishment) and measurable categorical variables (business category, e-catalog status) were further processed at the segmentation stage using the K-Means Clustering algorithm. Identity variables such as MSME names only served as markers of results, not as part of the analysis process.

3. **Clustering Stage**

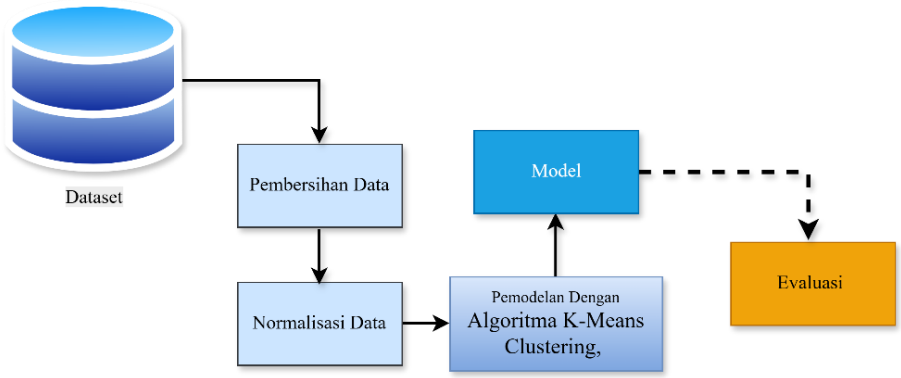

Figure 2. Segmentasi Flow

Data Pre-procesing

Data pre-processing is a critical stage in preparing the dataset for processing in the clustering algorithm. This stage includes data cleaning and normalization to ensure data quality, consistency, and readiness for further analysis. This process is important to reduce bias, improve accuracy, and ensure the validity of the segmentation Data cleaning is performed by identifying and handling missing values, duplicate data, and value inconsistencies. Incomplete data is corrected using simple imputation techniques or deleted if irrelevant, while duplicate data is eliminated to avoid bias in the analysis.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 1-13
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.760

Data normalization is applied using the Min-Max Normalization technique to standardize the scale of numerical variables into the range [0,1]. This normalization is necessary so that no variable dominates the Euclidean distance calculation in the K-Means Clustering algorithm, making the segmentation results more accurate and reliable.

Clustering

Clustering is the core of this research by applying the K-Means Clustering algorithm to group MSME data in Pekanbaru City based on certain characteristics. This algorithm was chosen for its efficiency, simplicity of implementation, and effectiveness in large-scale data segmentation [1,2,3]. Modeling was performed using Orange Data Mining software, which provides a visual interface and automation of the clustering iteration process.

The K-Means modeling process in Orange was carried out through the following stages:

1. Determining the K Value (Number of Clusters): The number of clusters was not determined a priori, but was explored by trying several k values (from 2 to 6). Each configuration is then evaluated using the Silhouette Score metric to determine the most optimal number of clusters.
2. Setting K-Means Parameters in Orange: Algorithm parameters are set through the Orange interface, including the number of clusters (k), maximum iteration limit, and centroid initialization method. The process of calculating Euclidean distances, assigning data to clusters, and updating centroids is performed automatically by the system.
3. Running the Clustering Process: Orange calculates the Euclidean distances between data points and groups MSMEs into clusters based on the similarity of the variables used (turnover, number of employees, length of establishment, business category, and e-Catalog participation status)
4. Recognizing the Cluster Patterns Formed: The modeling results are visualized in the form of a Scatter Plot and Silhouette Plot to understand the distribution and characteristics of each cluster, as well as the quality of separation between clusters.
5. Determining the Best Cluster: The configuration with the highest Silhouette Score value is selected as the final model. A value ≥ 0.5 is considered to indicate good cluster quality and clear separation between clusters.

Through this approach, MSME segmentation can be carried out systematically, measurably, and validly, resulting in homogeneous groups of MSMEs for more targeted policy analysis.

4. **Evaluation Stage**

The evaluation stage is conducted to assess the accuracy and validity of the data clustering results in representing the actual MSME segmentation. This evaluation is important to ensure that each cluster formed truly reflects the characteristics and level of MSME participation in the local e-Catalog, so that the results can be used as a basis for reliable analysis.

To measure the quality of clustering, this study uses the Silhouette Score metric. This metric works by comparing how close a piece of data is to other data in the same cluster (cohesion) and how far that data is from data in other clusters (separation). The Silhouette value ranges from -1 to 1. A value close to 1 indicates that the data has been grouped very well, with members in one cluster being very similar and clearly different from members of other clusters. A value around 0 indicates overlap between clusters, while a negative value indicates that the data may have been placed in the wrong cluster.

In this study, the clustering results are considered valid if the Silhouette Score reaches a minimum of 0.5. This threshold was chosen because it indicates that the clusters have been formed homogeneously and are clearly separated from other clusters. Using this criterion, the assessment of the segmentation results is not only quantitative but also allows for a more meaningful interpretation in the context of MSME characteristics. Through this evaluation, it is hoped that valid information can be obtained regarding the optimal number of clusters that match the data pattern, the quality of separation between clusters, and the level of uniformity of members within each cluster. Thus, the evaluation stage plays a key role in ensuring the reliability of the segmentation results before conducting a more in-depth analysis and formulating targeted policy recommendations.

2.3 Tools and Software

The main software used is Orange Data Mining 3, which is based on visual workflow. Orange facilitates all stages of analysis, from preprocessing (feature selection, imputation, and normalization), clustering modeling (using the K-Means widget), model evaluation (Silhouette Score), to visualization of results (Scatter Plot). In addition, Microsoft Excel is used for initial dataset cleaning and data format conversion (.CSV).

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 1-13
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.760

The research material consists of the 2024 MSME dataset obtained from the Pekanbaru City Cooperative and MSME Office and participation data from the Pekanbaru LPSE. This dataset includes key variables such as turnover, number of employees, length of establishment, business category, and E-Catalog participation status.

2.4 Research Location and Time

This research focuses on data objects in the Pekanbaru City area, with reference to secondary data for the 2024 Fiscal Year sourced from the Cooperative and MSME Office and the Pekanbaru City LPSE. The entire research process, from data acquisition to model evaluation, was carried out intensively from November to December 2025..

3. RESULTS AND DISCUSSION

3.1. Overview of Data

The research data covers a population of 10,560 verified MSMEs in the Pekanbaru City area in 2024. This dataset is the result of integration from three official government sources: the Cooperative and MSME Office (basic profile), LKPP, and the Inaproc v6 E-Catalog Portal (capacity and participation status data). The main variables analyzed consist of numerical data (monthly turnover, number of employees, length of business establishment) and categorical data (E-Catalog participation status) that have undergone a format alignment process for clustering analysis purposes.

3.2. Data Collection From Institutions

This study uses data obtained from three official institutions, namely the Pekanbaru City Cooperative and MSME Office, LKPP, and the e-Catalog v6 platform (inaproc.id). These three sources provide complementary information needed to build an integrated dataset as the basis for cluster analysis.

a. Data Sources and Formats

1. Pekanbaru City Cooperative and MSME Office

The Cooperative Office provides data on the verified MSME population in 2024, consisting of 10,560 units, including business names, addresses, business categories, legal status, and year of establishment. This data serves as the basic identity of MSMEs prior to the integration process.

Table 2. MSME in Pekanbaru City

Data Type	Description
Total Verified MSMEs (2024)	10,560 MSME units
Business Categories	Micro, Small, and Medium Enterprises
Business Profile Information	Business name, address, and type of business
Legal Status	Officially registered and verified
Data Year	2024

Table 3. Sample MSME in Pekanbaru City

Busines Name	Owner Name	Business Address	District	Business Category	NIB Status	Year Established
Warung Sari Rasa	Lina Marlina	Jl. Kenanga No.12	Marpoyan Damai	Kuliner	Ada	2018
Batik Lestari Riau	Anita Putri	Jl. Melur No. 7	Sukajadi	Kerajinan	Ada	2016
Toko Sinar Elektronik	Irwan Saputra	Jl. Nangka No. 21	Tampan	Elektronik	Tidak Ada	2020
Roti Manis Bakari	Yudi Pratama	Jl. Dahlia No. 8	Bukit Raya	Kuliner	Ada	2017
Dapur Kue Inara	Inara Wulandari	Jl. Cempaka No. 3	Lima Puluh	Kuliner	Tidak Ada	2021

Tables 2 and 3 are raw datasets obtained from the Pekanbaru City Cooperative and MSME Office. These two tables contain basic information on verified MSMEs in 2024, including business name, owner, address and sub-district, business category, legal status (NIB), and year of establishment. The data is still displayed in its original format as provided by the agency, so the variable structure has not been adjusted to the data from LKPP and the Local e-Catalog. Therefore, this table is used as an initial overview of the administrative characteristics of MSMEs as well as a basis for matching business identities in the process of cross-agency data integration.

2. Pekanbaru City Goods and Services Section

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 1-13
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.760

LKPP data includes the status of MSME participation in the e-Catalog (registered/unregistered) and the year of joining. This information is used as an indicator of the level of digital adoption by MSMEs and is an important variable in assessing the readiness of MSMEs to participate in the electronic procurement system.

Table 4. Participation Data for Pekanbaru City

Data Type	Description
MSME Participation in the Local E-Catalog	Status: registered / not registered
Data Period	2024

Table 5. Sample Data Participation E-Catalog Pekanbaru City

Business Name	E-Catalog Participation	Year Joined
Warung Sari Rasa	Registered	2021
Batik Lestari Riau	Registered	2020
Dapur Kue Inara	Not Registered	–
Roti Manis Bakari	Registered	2022
Craftwood Pekanbaru	Registered	2021

Tables 3.3 and 3.4 contain data on MSME participation in the Local e-Catalog obtained from LKPP. The information displayed includes the registration status of MSMEs in the e-Catalog and the first year they joined as providers. Each entry in the table represents the actual condition of digital adoption by each MSME, thereby illustrating their level of readiness to participate in the electronic-based government procurement system. This data is an important variable in the analysis because differences in participation status have the potential to affect the clustering results, where registered MSMEs generally have more developed administrative and digital capacities. Thus, the two tables are raw datasets from LKPP prior to integration with other data sources.

3. e-Catalog V6 Data (inaproc.id)

The e-Catalog platform provides quantitative data related to business capacity, namely the number of employees, annual/monthly turnover, length of time the business has been established, and type of product. These variables are key components in cluster formation because they describe the economic scale and operational capabilities of MSMEs.

Table 6. V6 Pekanbaru City E-Catalog Data

Data Type	Description
Monthly MSME Revenue	Within the range defined by national MSME thresholds (≤ IDR 500 million per month)
Number of Employees	Total number of active employees in each MSME
Business Age	Year of establishment – 2024
Product Type	Category of goods or services offered
Data Validity	Data obtained directly from the official registered MSME pages

Table 7. Sample Data E-Catalog V6 Pekanbaru City

Business Name	Number of Employees	Business Class	Annual Revenue (IDR)
Warung Sari Rasa	4	Micro	120,000,000
Batik Lestari Riau	6	Small	350,000,000
Toko Sinar Elektronik	3	Micro	95,000,000
Roti Manis Bakari	5	Micro	180,000,000
Craftwood Pekanbaru	8	Small	420,000,000

Tables 3.5 and 3.6, sourced from the Local e-Catalog platform (inaproc.id v6), contain key quantitative data used in cluster analysis, including the number of workers, business class, and annual turnover of MSMEs that have been

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 1-13
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.760

registered as official providers. The information in these two tables illustrates the operational capacity of each business actor, which can be used to assess business scale, productivity levels, and growth potential. These variables are key components in cluster formation because they show significant variations in the economic scale and production capacity of MSMEs. The datasets in Tables 3.5 and 3.6 are still raw data before integration with sources from the Cooperative and MSME Office and LKPP, so they serve as a starting point in the process of variable normalization and MSME grouping in the next stage of analysis.

b. Data Integration

Data from the Cooperative and MSME Office, LKPP, and the Local e-Catalog platform (inaproc.id v6) have different structures and formats, requiring an integration process to form a consistent dataset. Initial alignment, including column name normalization, data type adjustment, and administrative element cleaning, was performed using Microsoft Excel. Further integration was then carried out using Python (pandas) through merge and join operations.

The integration process uses the three most stable identity attributes across all data sources, namely business name, address/district, and business category. These three attributes serve as the key to matching between datasets so that business capacity variables (turnover, workforce, length of establishment) and digitization variables (e-Catalog participation) can be accurately combined. An example of the integration results is shown in the following table:

Table 8. Data Sample After Integration

No.	MSME Name	MSME Category	Number of Employees	Monthly Revenue (IDR)	Business Age (Years)	E-Catalog Participation
1	MSME_3282	Fashion	71	IDR 25,500,000	1	1
2	MSME_3536	Fashion	64	IDR 46,400,000	18	1
3	MSME_4188	Fashion	15	IDR 2,600,000	12	0
4	MSME_8707	Handicrafts	21	IDR 550,000	16	1
5	MSME_1114	Fashion	68	IDR 2,600,000	18	0
6	MSME_1663	Fashion	8	IDR 45,000,000	4	0
7	MSME_7051	Food & Beverage	90	IDR 3,000,000	18	0
8	MSME_10122	Food & Beverage	73	IDR 2,800,000	15	1
9	MSME_3778	Food & Beverage	98	IDR 4,700,000	18	0
10	MSME_10072	Handicrafts	71	IDR 3,300,000	1	1

This final dataset is then used as the basis for analysis in the next stage.

c. Variable Identification

Research variables are divided into three categories: business identity (name, category, location), business capacity (turnover, workforce, length of establishment), and digitization (e-Catalog participation). The numerical variables of turnover, workforce, and length of establishment form the core of the cluster analysis, while the categorical variables support the interpretation of the results. The selection of these variables ensures that the segmentation of MSMEs comprehensively reflects economic, operational, and digital readiness conditions. The following are the variables used:

Table 9. Data variables

Variable	Data Source	Data Type	Variable Role Description
Business Name	Department of Cooperatives, LKPP, E-Catalog	Categorical (String)	Used as a primary identifier for data matching during the data integration process; not included in the clustering analysis.
Business Category	Department of Cooperatives	Categorical	Indicates the type of MSME commodity; serves as descriptive information for the study.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 1-13
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.760

Year of Establishment	Department of Cooperatives & E-Catalog	Numerical (Integer)	Used to calculate business age, which is subsequently processed as a quantitative variable.
Local E-Catalog Participation	LKPP	Categorical (Binary: Registered / Not Registered)	Indicator of MSME digitalization level; converted into numerical form during the preprocessing stage.
Annual Revenue	E-Catalog v6	Numerical (IDR)	Core variable in cluster formation as it reflects the economic scale of the MSMEs.
Number of Employees	E-Catalog v6	Numerical	Key variable for clustering, representing the MSME's workforce capacity.

These variables were selected because they provide a comprehensive picture of the condition of MSMEs, in terms of administrative identity, production capacity, and level of digital adoption.

3.3. Segmentation

MSME segmentation was carried out through two main steps: (1) data preprocessing and (2) the application of three clustering algorithms, namely K-Means, DBSCAN, and Hierarchical Clustering, using an integrated dataset of 10,560 MSMEs.

1. Preprocessing

Data preprocessing using the Preprocess widget in Orange Data Mining includes imputation of missing values (numeric filled with the average, categorical filled with the mode), normalization of numeric variables (turnover, workforce, length of establishment) to the interval [0,1], and transformation of categorical variables (MSME category and e-Catalog participation) into dummy variables (one-hot encoding) to be ready for use in distance-based clustering algorithms. Here are the results:

Table 10. Data variables

MSME Name	MSME Category	Number of Employees (Normalized)	Business Age (Years, Normalized)	Monthly Revenue (IDR, Normalized)	Participation_0	Participation_1
UMKM_1	Makanan	0.6020	0.4211	0.0396	1	0
UMKM_2	Makanan	0.1020	0.8421	0.0155	1	0
UMKM_3	Perdagangan	0.4490	0.7895	0.0448	0	1
UMKM_4	Makanan	0.9796	0.5263	0.0504	1	0
UMKM_5	Jasa	0.5000	0.0526	0.0065	1	0
UMKM_6	Jasa	0.8367	0.2632	0.1445	0	1
UMKM_7	Jasa	0.7857	0.2105	0.0905	1	0
UMKM_8	Makanan	0.6531	0.9474	0.0205	1	0
UMKM_9	Jasa	0.9694	0.7895	0.0329	0	1
UMKM_10	Fashion	0.4489	0.6842	0.0092	1	0
UMKM_11	Makanan	0.6633	10.000	0.0442	1	0
UMKM_12	Kerajinan	0.9796	0.2632	0.0389	0	1
UMKM_13	Fashion	0.7245	0.2105	0.0174	1	0

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 1-13
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.760

UMKM_14	Kerajinan	0.5102	0.6842	0.0025	1	0
UMKM_15	Kerajinan	0.1327	0.1579	0.2318	0	1
UMKM_16	Jasa	0.2857	0.5263	0.0021	1	0
UMKM_17	Jasa	0.7857	0.3684	0.0002	0	1
UMKM_18	Kerajinan	0.9388	0.7368	0.0056	1	0
UMKM_19	Fashion	0.5510	0.5263	0.0231	0	1
UMKM_20	Jasa	0.6020	0.3158	0.0027	0	1

After preprocessing, the dataset is ready to be analyzed with five numeric variables as the main inputs. Additional variables such as name and MSME category are retained as metaattributes for interpreting cluster results, without affecting distance calculations.

The table shows the data structure after preprocessing, where the numerical variables (Number_of_Employees, Years_in_Business, Monthly_Turnover) have been normalized to a range of 0–1, while the columns Participation=0 and Participation=1 are the results of one-hot encoding of categorical variables, with a value of "1" indicating the MSME's participation status in the e-Catalog. Nama_UMKM and Kategori_UMKM are retained as meta attributes for interpreting the results, without affecting the cluster calculation.

2. Clustering

After preprocessing, K-Means Clustering is applied to group MSMEs based on four normalized numeric variables: Number of Employees, Length of Business Establishment, Monthly Turnover, and e-Catalog Participation. The goal is to identify groups of MSMEs with similar characteristics related to business capacity and participation in the e-Catalog. The number of clusters is determined based on the data distribution pattern on the Scatter Plot in Orange, resulting in k = 3 as the representative configuration. The algorithm calculates the distance of each MSME to the centroid, updating the centroid position until convergence. The results divide MSMEs into three clusters: C1 (low turnover, low participation), C2 (medium characteristics), and C3 (large business scale, active participation). The following table shows a snapshot of the clustering results from Orange.

Table 11. Modeling results data

MSME Name	MSME Category	Cluster	Number of Employees	Business Age	Monthly Revenue	E-Catalog Participation	r_participation
UMKM_1	Makanan	C3	0.6020	0.4211	0.0396463	0	1
UMKM_2	Makanan	C3	0.1020	0.8421	0.0155675	0	1
UMKM_3	Perdagangan	C2	0.4490	0.7895	0.0447969	1	0
UMKM_4	Makanan	C1	0.9796	0.5263	0.906449	0	1
UMKM_5	Jasa	C2	0.5000	0.0526	0.00568386	1	0
UMKM_6	Jasa	C2	0.8367	0.2632	0.144573	1	0
UMKM_7	Jasa	C1	0.7653	0.2105	0.808808	1	0
UMKM_8	Makanan	C2	0.6531	0.9474	0.0240543	0	1
UMKM_9	Jasa	C2	0.9694	0.7895	0.0392953	0	1
UMKM_10	Fashion	C3	0.4898	0.8947	0.00239453	0	1

This table shows the distribution of MSMEs into three K-Means clusters based on Number of Employees, Monthly Turnover, Length of Operation, and e-Catalog Participation. Cluster C1 consists of micro-scale MSMEs with low turnover, few employees, and minimal digital participation, thus requiring intensive assistance. C2 includes developing MSMEs with medium capacity and a tendency to participate in e-Catalogs, showing potential for improvement if supported. Meanwhile, C3 consists of established MSMEs with high turnover, larger workforce, and active participation, reflecting optimal readiness for digital integration and utilization of e-Catalog services. These three clusters provide an overview of MSME segmentation based on business capacity and level of digital adoption.

3.4. Evaluation Using Silhouette Score

The quality of clustering is evaluated using the Silhouette Score, which measures the extent to which data is close to its own cluster compared to other nearby clusters. Silhouette values range from -1 to +1, where values close

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 1-13
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.760

to +1 indicate well-formed clusters, values around 0 indicate overlap between clusters, and negative values indicate incorrect grouping.

During the evaluation stage, the K-Means algorithm was tested with varying numbers of clusters: K = 2, 3, 4, and 5. Each configuration produced a different Silhouette Score value, allowing for a comparison of cluster formation quality. The following table presents the Silhouette Score results from K-Means Clustering.

Table 12. Modeling results data

Number of Clusters (K)	Silhouette Score
2	0.417
3	**0.444**
4	0.361
5	0.323

The evaluation results show that the highest Silhouette value was obtained at K = 3, namely 0.444, compared to K = 2 (0.417), K = 4 (0.361), and K = 5 (0.323). This value indicates that increasing the number of clusters above 3 reduces the quality of separation because the clusters overlap more, while K = 2 produces segmentation that is too broad and less representative. The selection of K = 3 is considered optimal because it provides the best balance between inter-cluster distance and internal variance, and is in line with the characteristics of MSME business capacity: micro, medium, and relatively large. Thus, the three-cluster K-Means is the most representative configuration for the segmentation of MSMEs in Pekanbaru City.

After that, a Silhouette Plot visualization was added to reinforce the evaluation results. The graph shows that the cluster with a value of K = 3 has the most stable curve shape and the clearest separation between clusters compared to other configurations. This is in line with the highest Silhouette value obtained, so that visually and numerically, three clusters are the most optimal grouping structure.

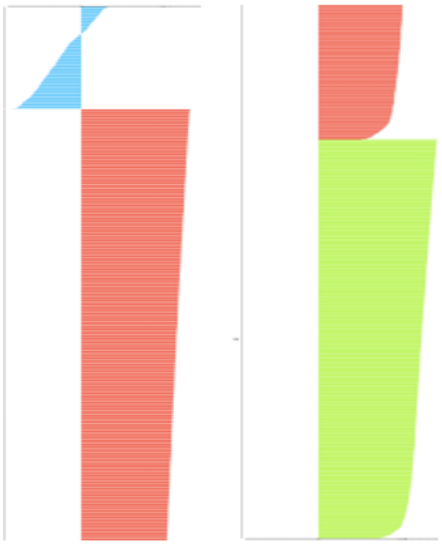

Figure 3. Silhouette score visualization

The silhouette plot visualization shows differences in clustering quality between clusters. The blue block appears open in both directions because the silhouette values in this cluster are more varied; some data has good cluster matching, while others are closer to other clusters, resulting in low or negative values. Meanwhile, the red and green blocks appear longer and thicker because the number of MSMEs in both clusters is larger and their silhouette values are more stable and consistent in one direction. This shows that the red and green clusters are more solid than the blue cluster.

3.5. Discussion of Cluster Results

The K-Means algorithm provides stable segmentation with three clusters that have clear economic meanings for small, medium, and relatively large business groups. The Silhouette value of 0.444 indicates good separation quality, with relatively low internal variance and a sufficiently large margin between clusters. The symmetrical cluster

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 1-13
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.760

structure makes K-Means very suitable for large normalised datasets. K-Means is able to describe the diversity of business capacity while directly linking it to the pattern of participation in local e-catalogues as follows:

Table 13. Discussion of cluster results

Cluster	Key Characteristics	Interpretation
C1	Low revenue, limited workforce, minimal participation	Micro-scale MSMEs requiring digital capacity building
C2	Medium business scale, stable performance, increasing capacity	Growing MSMEs with potential for higher digital participation
C3	High revenue, large workforce, active participation	Established MSMEs ready for full integration into the e-catalog system

This table summarizes the main characteristics of the three K-Means clusters: C1, C2, and C3. Cluster C1 includes MSMEs with the lowest business capacity, small turnover, limited workforce, and minimal participation in e-catalogues, thus requiring digital coaching. Cluster C2 contains medium MSMEs with stable capacity and increasing participation, making them a potential group for strengthening digital literacy. Cluster C3 consists of MSMEs with high performance, large turnover and workforce, and active participation in e-catalogues, ready to fully adopt digital procurement systems. These results show that the greater the business capacity, the higher the tendency to participate in e-catalogue

4. CONCLUSION

This study successfully segmented SMEs in Pekanbaru City based on business capacity and level of participation in the Local E-Catalog by utilizing three clustering algorithms, namely K-Means.The integration of data from the Cooperative and SME Office, LKPP, and the E-Catalog v6 platform produced an integrated dataset that can be used for effective clustering. Based on the evaluation results using the Silhouette Score, K-Means with three clusters provided the best performance (0.444), indicating the most stable cluster separation compared to other configurations. This segmentation resulted in three MSME groups: (1) small MSMEs with low business capacity and minimal digital participation, (2) developing MSMEs with medium capacity and potential for increased digital adoption, and (3) established MSMEs with high turnover and active participation in the e-Catalog. These findings show that business capacity has a direct relationship with the level of MSME involvement in the government's digital procurement ecosystem. The results of this study suggest that local governments use SME segmentation as a basis for more targeted guidance, particularly in improving the digital capabilities of low-capacity SMEs. SMEs also need to strengthen their technological literacy in order to be better prepared to utilize the e-Catalog. For further research, it is recommended to add other variables or clustering methods to produce a more comprehensive segmentation.

REFERENCES

[1] E. Syarifah, S. Purnamasari, And A. Purnomo, “Efektivitas Penyaluran Dana Banpres Produktif Usaha Mikro (Bpum) Untuk Modal Kerja Dalam Meningkatkan Kesejahteraan Pelaku Umkm,” 2020.
[2] T. Jelita, R. Buaton, And M. Simajuntak, “Pengelompokan Bidang Usaha Terhadap Bantuan Produktif Usaha Mikro (Bpum) Berdasarkan Wilayah Deli Serdang Menggunakan Metode Clustering K-Means (Studi Kasus: Dinas Koperasi Dan Umkm Kabupaten Deli Serdang),” *J. Comput. Sci. Inf. Technol.*, July 2023.
[3] D. T. Alamanda, E. Kusmiati, D. F. Shiddieq, And F. F. Roji, “Msme Clusterization Using K-Means Clustering In Garut Regency, Indonesia,” 2023.
[4] A. Maahira, “Analisis Program Pemberdayaan Masyarakat E-Katalog Umkm Untuk Memajukan Ekonomi Masyarakat Kota Medan,” *J. Obor Penmas Pendidik. Luar Sekol.*, Vol. 6, No. 1, Pp. 51–60, Apr. 2023, Doi: 10.32832/Oborpenmas.V6i1.14356.
[5] M. R. Fahlevi, D. R. D. Putri, And E. Syahrin, “Analisis Pengelompokan Data Pelelangan Barang Dengan Metode K-Means Clustering,” Vol. 8, 2023.
[6] W. Budiono, B. I. Nugroho, And N. A. Santoso, “Penerapan Metode Fuzzy K-Means Clustering Untuk Pengelompokan Konten Halaman Web Secara Otomatis,” 2024.
[7] I. Iin, R. Fadila, A. Rizki Rinaldi, And F. Fathurrohman, “Penerapan Data Mining Dalam Mengelompokan Jumlah Umkm Berdasarkan Kabupaten Kota Menggunakan K-Means Clustering,” *Jati J. Mhs. Tek. Inform.*, Vol. 8, No. 2, Pp. 1446–1450, Apr. 2024, Doi: 10.36040/Jati.V8i2.8427.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 1-13
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.760

[8] H. Mawarni, G. Testiana, And M. L. Dalafranka, "Implementasi Algoritma K-Means Untuk Segmentasi Pelanggan Pada Pt. Bintang Multi Sarana Cabang Tugumulyo," *J. Komput. Dan Inform.*, Vol. 11, No. 2, Pp. 227–236, Oct. 2023, Doi: 10.35508/Jicon.V11i2.12478.
[9] F. Alzami *Et Al.*, "Implementation Of Etl E-Commerce For Customer Clustering Using Rfm And K-Means Clustering," *J. Ilm. Merpati Menara Penelit. Akad. Teknol. Inf.*, Vol. 10, No. 3, P. 167, Dec. 2022, Doi: 10.24843/Jim.2022.V10.I03.P05.
[10] A. Azzam, A. Irma Purnamasari, And I. Ali, "Implementasi Algoritma K-Means Clustering Untuk Analisis Persebaran Umkm Di Jawa Barat," *Jati J. Mhs. Tek. Inform.*, Vol. 8, No. 3, Pp. 3062–3070, May 2024, Doi: 10.36040/Jati.V8i3.8450.
[11] M. Afrizal, I. Saputra, And R. Satria, "Analisis Performa Algoritma K-Means Clustering Untuk Segmentasi Pasar Di Umkm," Vol. 5, No. 2, 2025.
[12] A. Ahmad, "Implementasi E Katalog Terhadap Perkembangan Umkm Di Kota Pekanbaru," Vol. 6, 2024.
[13] Sinta Puspita Sari, Amyra Syalsabila, Silvi Yulianti, And Sigit Djalu Purwoko, "Studi Literatur Cara Menjadi Penyedia E-Katalog Pada Lkpp Sebagai Lembaga Pengadaan Barang Dan Jasa Bagi Pelaku Usaha," *J. Manuhara Pus. Penelit. Ilmu Manaj. Dan Bisnis*, Vol. 2, No. 2, Pp. 64–71, Jan. 2024, Doi: 10.61132/Manuhara.V2i2.706.
[14] A. A. Khan, M. S. Bashir, A. Batool, M. S. Raza, And M. A. Bashir, "K-Means Centroids Initialization Based On Differentiation Between Instances Attributes," *Int. J. Intell. Syst.*, Vol. 2024, No. 1, P. 7086878, Jan. 2024, Doi: 10.1155/2024/7086878.
[15] W. T. Pambudi And A. Witanti, "Penerapan Algoritma K-Means Clustering Untuk Menganalisis Penjualan Pada Toko Ayu Collection Barbasis Web," Vol. 6, No. 3, 2021.
[16] N. A. Jauza And M. Albina, "Model Dan Pendekatan Penelitian Kuantitatif: Kajian Filosofis, Metodologis, Dan Aplikatif," Vol. 2, 2025.