International Journal of Entrepreneurship and Sustainability Studies, Vol. 5 No. 2 (2025) https://doi.org/10.31098/ijeass.v5i2.3419

Check for updates

Research Paper

Regional Segmentation Based on the Level of Digitalization of MSMEs in Indonesia

Antradiva Oktaviola Puspita, Dynda Aisyah Dyah Pitaloka, Deky Aji Suseno
Universitas Negeri Semarang, Indonesia

Received : June 11, 2025	Revised : July 7, 2025	Accepted : November 21, 2025	Online : December 12, 2025

Abstract

Digitalization of Micro, Small, and Medium Enterprises (MSMEs) plays a vital role in enhancing national and international competitiveness. This study aims to conduct regional segmentation in 34 provinces in Indonesia based on several indicators of the level of digitalization of MSMEs. This research is a quantitative study and an analytical descriptive approach that uses analysis techniques (K-means cluster) to map the distribution of MSMEs in Indonesia. Data pre-processing by standardizing data (z-score) to overcome differences in units of measurement. The type of data used is secondary data from the Central Bureau of Statistics and the Indonesian Payment System Association. The study identifies three clusters with different characteristics. The first cluster, comprising West Java, Central Java, and East Java, shows advanced MSMEs digitalization. The second cluster consists of 30 provinces with a developing level of MSMEs digitization. The third cluster is occupied by Papua, which shows lagging MSMEs' digitalization. This research focuses on the uneven digitization of MSMEs by limiting the scope of research to 34 provinces in Indonesia, then does not include analysis of development in time series, financial aspects, or the impact of government policies on the development of MSMEs. The originality of this research lies in the use of 34 research objects in Indonesia in 2023 by applying non-hierarchical K-Means clustering analysis. The analysis in this study uses indicators including the distribution of MSMEs, the number of workers in the e-commerce sector, the number of e-commerce businesses, the level of QRIS adoption, and the use of social media.

Keywords: *K-Means Clustering, MSMEs, Digitalization.*

INTRODUCTION

Digital transformation has changed the global economic landscape, creating opportunities for various sectors, including Micro, Small, and Medium Enterprises (MSMEs). As the main pillar of the Indonesian economy, MSMEs have a strategic role in driving national economic growth. Data from the Central Statistics Agency (BPS) in 2023 recorded a positive growth in the number of MSMEs with an increase of 162,320 units compared to 4,339,228 units in the previous year. The utilization of digital technology allows MSME players to expand market access, improve operational efficiency, and strengthen competitiveness amidst changing economic dynamics (Basri et al., 2023).

Digitalization of MSMEs not only reduces financial barriers but also expands and diversifies the supply chain. This provides greater space for businesses to optimally allocate human resources and encourage the utilization of technological innovation in their business development. However, the implementation of digitalization among Indonesian MSMEs still faces various obstacles that affect productivity, where MSMEs only reach around 40% compared to large companies (Ayu Nursasi et al., 2024). The delay in the spread of MSMEs' digitalization is not solely caused by infrastructure limitations, but also by internal factors such as the adoption rate of a workforce that has an understanding of the digitalization process, the ability to adapt to new technologies, and the effectiveness of using social media as a modern marketing channel (ILO & OECD, 2022).

Corresponding author's email: antradiva994@students.unnes.ac.id

The issue of MSMEs' digitization becomes more complex when viewed from a geographical perspective, where disparities in access to technology between regions are a major challenge that must be faced. This disparity is evident from the low number of e-commerce units in some regions, such as East Nusa Tenggara and Papua, which only reach 694,673 and 748,109 units, respectively. This disparity is further reinforced by the data on social media adoption, where, despite a national increase of 0.55% in social media usage, the two provinces recorded the lowest adoption rates, at 78.88% for East Nusa Tenggara and 38.71% for Papua. This regional imbalance is also revealed in the employment of the digital sector, which is still concentrated in regions with better infrastructure. Java Island dominates employment in the e-commerce sector with a contribution of 79.52% of the total national workforce, while other regions are left behind due to limited digital infrastructure. This condition shows that the great potential of digitalization in creating job opportunities has not been able to be utilized evenly throughout Indonesia.

Facing the challenge of regional disparities in MSMEs' digitalization, the Indonesian government has implemented various strategic policies and programs. The government responded to the persisting digital divide by launching the Digital Technology Adoption and Business Acceleration for MSMEs 2024 program, which is focused on addressing digital inequality in various regions of Indonesia (Wijayantini et al., 2024). One of its main initiatives, the Level Up program, is aimed at accelerating digital transformation in the MSMEs sector with a target of 40 million digitized MSMEs by 2024. However, the implementation of these programs is not fully in line with empirical findings that emphasize the importance of accuracy in digital reporting, especially in terms of financial recording. The study results show that digital-based financial records can accelerate the growth rate of MSMEs, but the distribution of digitalization does not always touch the technical aspects of operations equally.

Research on the digitalization of MSMEs has developed with a focus on various aspects of the implementation of digital technology in small and medium enterprises. Handini and Choiriyati (2019) examined the digitalization of MSMEs as a result of innovations in marketing communications, particularly during the COVID-19 pandemic, which shows the importance of digital adaptation in crises. From a financial perspective, Reski (2024) examined the importance of preparing MSMEs' financial reports, while Ayem and Wahidah (2021) analyzed factors affecting the financial performance of MSMEs in Yogyakarta City. Research exploring the benefits of digitizing MSMEs' financial records has been conducted with a focus on efficiency and effectiveness (Pakpahan, 2021).

Although various studies have examined MSMEs' digitalization, there are still research gaps that need to be noted. First, the majority of previous studies have focused on the technical and operational aspects of digitalization, but have not openly analyzed regional disparities in the implementation of MSMEs' digitalization in Indonesia. Second, no study has used a clustering approach to identify and group regions based on the level of MSMEs digitalization. Therefore, this study aims to identify regions that have not met the indicators of digitalization by examining some key variables such as the distribution of MSMEs, the number of e-commerce players, the workforce in the digital sector, the use of QRIS, and the intensity of social media use in business activities by clustering provinces in Indonesia through the K-Means Clustering method approach, where the final result is a map visualization as a reference in the formulation of region-based MSMEs development policies.

LITERATURE REVIEW

Theoretical Framework: Innovation Diffusion Theory

This study adopts Rogers' (2024) Innovation Diffusion Theory as the theoretical foundation to understand the uneven adoption of digital technologies among MSMEs across Indonesian provinces. The theory explains how innovations spread through social systems over time, influenced by five key characteristics: relative advantage, compatibility, complexity, trialability, and observability (Rogers, 2024). In the context of MSME digitalization, relative advantage refers to the perceived benefits of digital technologies such as operational efficiency, cost reduction, and market expansion. Compatibility relates to how well digital solutions align with existing business practices and cultural values. Complexity involves the technical difficulty MSMEs face in implementing digital technologies. Trialability allows MSMEs to experiment with digital solutions before full adoption, while observability enables them to witness the results of digital adoption in similar businesses.

This theoretical framework helps explain why certain provinces demonstrate advanced digitalization patterns (early adopters) while others exhibit lagging adoption (laggards). The theory suggests that innovation adoption follows a predictable pattern, creating natural clusters of adopters that can be identified through empirical analysis (Venkatesh & Davis, 2000). This provides the conceptual foundation for regional segmentation based on digitalization levels.

MSMEs Digitalization and Development

Technological advances demand the implementation of digitalization in various sectors of life, including in the economy. Micro, Small, and Medium Enterprises (MSMEs) are one of the sectors that have undergone the digital transformation process. MSMEs that adopt digitalization tend to have higher competitiveness and are able to survive longer in the face of market dynamics. Research by Rusdiaman et al (2024) shows that the sustainability of MSMEs is highly dependent on the integration of digitalization in their business operations. The inability to utilize digital technology, especially e-commerce platforms, can harm business performance, such as sales stagnation to the risk of business closure. The low level of utilization of digitalization by MSME players is caused by several factors, including limited digital literacy, especially in terms of technical management of online stores, for example, difficulties in opening a store on an e-commerce platform, and ignorance in optimizing available features. In addition, concerns regarding the high shipping costs that must be borne by consumers are also an obstacle, especially for business actors (Kurnia & Wulandari, 2022).

Clustering Analysis in MSME Research

With similar problems, previous research also strengthened these findings through the analysis of digitalization indicators that significantly affect the performance of MSMEs. This research was then complemented and expanded by a follow-up study that used a k-means classification approach, which openly evaluated three main variables: the percentage of internet users, the percentage of e-commerce users, and the proportion of MSMEs that have utilized the internet in their business operations (Idris, 2024). This approach provides a more detailed mapping of the main determinants of inefficiency, particularly in the real sector, and strengthens the argument that digitalization is a strategic factor in improving the competitiveness and sustainability of MSMEs.

Digital Technology Adoption Indicators

The measurement of MSME digitalization requires multiple indicators that capture different dimensions of digital adoption. E-commerce participation represents the most visible form of digital transformation, indicating MSMEs' ability to access online markets (Anatan & Nur, 2023). Social media utilization reflects digital marketing capabilities and customer engagement strategies (Drummond et al., 2020)

Digital payment systems, particularly QRIS adoption, represent fundamental infrastructure for digital transactions. Research by Tazkia et al. (2024) identifies four distinct clusters of QRIS users in Indonesia, highlighting the importance of payment digitalization in MSME operations. Workforce digitalization, measured through e-commerce sector employment, indicates the human capital dimension of digital transformation (Lv et al., 2025). The integration of these indicators provides a comprehensive view of regional digitalization levels, enabling more accurate segmentation and policy recommendations

RESEARCH METHOD

This research uses a quantitative approach to conduct an in-depth analysis of classification techniques. The classification approach used refers to a systematic process of grouping research objects into certain categories or classes that are arranged in a hierarchical and structured manner, as proposed by Montesi and Urdiciain (2005). In the implementation of this research, the clustering process utilizing the k-means algorithm resulted in three main groups representing the classes formed.

Table 1. Variable Description

Variable	Description	Unit
Number of MSMEs deployments	Total MSMEs that have adopted digital technology (including e-commerce, social media, and digital payments) in business activities.	Business unit
Social media users	The number of active social media user accounts used for business purposes or promotion of MSMEs' products.	People
Total e-commerce	The number of e-commerce platforms operating legally and actively in a region (national or regional), including marketplaces and standalone online stores.	Company
QRIS merchant users	Number of businesses officially registered as QRIS (Quick Response Code Indonesian Standard) users for digital transactions.	Business Unit
Labor in e-commerce	The number of individuals working in the e-commerce sector, both in e-commerce companies and digital MSMEs businesses.	People

Sources: Data Processed, 2025

Based on the operational definitions of the variables listed in Table 1, this study takes all provinces in Indonesia as the unit of analysis, namely 34 provinces before the latest administrative area expansion. The data used is data for 2023. This study combines a number of variables that represent indicators of digitalization, and then analyzes them using clustering techniques with the K-Means algorithm. The K-Means method was chosen because of its ability to group data into several homogeneous clusters based on certain characteristics. This clustering process is carried out by taking into account important stages, such as data preprocessing and transformation (Bock, 2008).

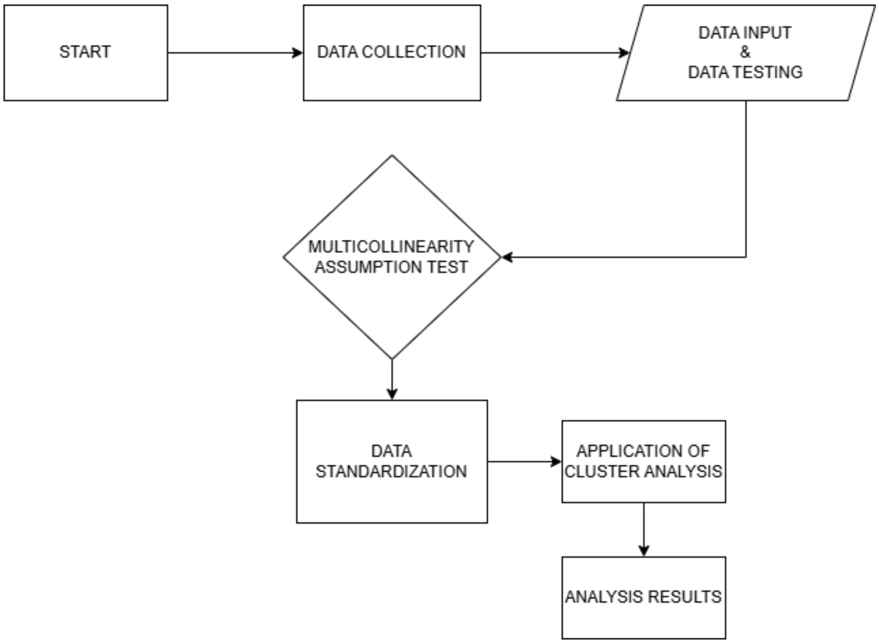

Figure 1. Flow of Clustering Application with K-Means Clustering Method
Source: Data Processed (2025)

The stages of the methodology that will be carried out in this research are described in a methodology flow as shown below:

Data Collection:

Data collection in this study was carried out using secondary data obtained through trusted sources. The method used is a documentation study, which is a data collection technique carried out through a review of documents and records that are already available. The main data sources used were the official websites of the Central Statistics Agency (2024) and the Association of Indonesian Planning Schools (2023). After the data was collected, the researcher conducted a data analysis and processing process to extract relevant information and support the objectives of this study.

Data Input & data testing

Once the data had been collected, the next step was to input the data into the analysis tool to undergo a series of statistical processing stages. At the initial stage, a multicollinearity test was conducted to identify the presence of high linear relationships between independent variables, which could compromise the validity of the analysis results. In addition, a cluster standardization test was conducted to ensure that all variables were on a comparable scale before the clustering method was applied.

Multicollinearity Assumption Test

The multicollinearity test is a procedure used to detect the correlation between independent variables in a cluster model. The purpose of this test is to ascertain whether there is interdependence or collinearity among the indicators of the research variables that can affect the validity of the model (Gujarati, 2015). The indicators used in this test are the Variance Inflation Factor (VIF) and tolerance values.

- If the VIF value > 10 or the tolerance value is below 0.1, then it indicates the presence of multicollinearity symptoms.
- If the VIF value is < 10 and the tolerance is more than 0.1, then the regression model can be said to be free from multicollinearity.

Data Standardization

Data standardization is the process of converting data to a common format to allow users to process and analyze it (Yadohisa et al., 2010). Data standardization is necessary so that all variables are on a comparable scale before analysis. Therefore, the data is converted into z-scores or other standardized scales to avoid bias in the analysis.

Application of Cluster Analysis

The K-Means Clustering algorithm aims to partition n observations into k clusters, where each observation is classified to the cluster with the closest mean (centroid). The following are the fundamental equations in K-Means Clustering:

- Calculation of Euclidean Distance between data point and centroid
- New Centroid Calculation
- Objective Function (Sum of Squared Error)

Data Mining

A series of processes to extract added value from a data set in the form of knowledge that was unknown manually. This methodology involves structured stages in exploring and analyzing data to find hidden patterns. This process transforms raw data into actionable information through a series of systematic stages (Verma et al., 2012)

A set of processes to extract value from a data set refers to a systematic approach that aims to identify hidden knowledge that cannot be discovered through manual means. This process is known as data mining, which is an integral part of the discipline of data science and artificial intelligence. The main objective of this methodology is to extract relevant, meaningful, and valuable information from raw data that initially appears unstructured or insignificant.
The methodology involves some structured stages, ranging from data preprocessing, data exploration, feature selection, to the application of analysis algorithms such as classification, clustering, association, and prediction (Kumar et al., 2020). In each stage, data is evaluated and processed in such a way as to identify hidden patterns, relationships between variables, or trends that can provide new insights. This stage is not only technical but also requires a contextual understanding of the data domain used, so that the analysis results obtained are truly relevant and can be applied in real life.

The transformation of raw data into actionable information is at the core of this process. The information can then be used in strategic decision-making, policy planning, or improving operational efficiency in various sectors, such as business, health, education, and government. In other words, this process not only acts as a technical tool but also as an important instrument in the development of data-based insights and innovations.

In the context of the industrial revolution 4.0 and the current era of big data, the utilization of data exploration methods like this is becoming increasingly crucial. The ability to understand and interpret large amounts of data in an efficient and accurate manner is a significant competitive advantage. Therefore, mastering this process is not only important for data practitioners but also for policy makers and industry players who want to optimize the potential of data as a strategic asset.

K-Means Clustering

K-Means Clustering is an unsupervised machine learning algorithm designed to partition n observations into k clusters, where each observation is classified into the cluster with the nearest centroid (mean) (Remawati et al., 2021; Eka et al., 2014).

K-Means Algorithm Steps

The K-Means algorithm iteratively optimizes centroid placement to minimize inertia, or the sum of squared distances between data points and their respective cluster centroids. The detailed steps are as follows:

Determining the Number of Clusters (k)

The initial step in implementing the K-Means algorithm involves determining the desired number of clusters, denoted as k. The value of k must be established before algorithm execution. The Elbow Method is a commonly used technique for determining the optimal k value.

Initializing Centroids

Once k is determined, k initial centroids (cluster centers) are randomly initialized within the data space. These initial centroids serve as the starting reference points for the data grouping process.

Data Grouping

In this step, each data point is assigned to the cluster whose centroid is closest. The commonly used distance function is the Euclidean Distance, defined as:

$$d\ (x, c) = \sqrt{\sum xi - ci}$$

where:

- d(x,c) is the distance between data point x and centroid c.
- xi is the i-th feature value of data point x.
- ci is the i-th feature value of centroid c.
- D is the dimensionality of the feature space.

This process involves calculating the distance from each data point to all existing centroids, then assigning that data point to the cluster with the nearest centroid.

Centroid Update

After all data points have been assigned to their respective clusters, the centroid positions for each cluster are updated. The new centroid is calculated as the arithmetic mean of all data points belonging to that cluster. The calculation of the new centroid (cj) for cluster j is formulated as:

$$cj = 1\ /\ |Sj| \sum xi\ \epsilon\ S_j X_i$$

where:

- cj is the new centroid for cluster j.
- Sj is the set of data points belonging to cluster j.
- |Sj| is the number of data points in cluster j.

- x is a data point in cluster j.

These updated centroids will serve as the reference points for the next iteration.

Iteration until Convergence

Steps 3 (Data Grouping) and 4 (Centroid Update) are repeated iteratively until a convergence criterion is met. Convergence is achieved when:

- Centroid positions no longer change significantly.
- Data point assignments to clusters remain unchanged.
- The objective function reaches its minimum value.

Objective Function (Sum of Squared Error)

The quality of clustering is evaluated using an objective function known as the Sum of Squared Error (SSE), formulated as:

$$J = \sum k_{j=1} \sum zi \in Sj \; \|Xi - Cj\|^2$$

where:

- J is the value of the objective function to be minimized.
- k is the number of clusters.
- x is a data point.
- cj is the centroid of cluster j.
- ||x−cj||2 is the squared Euclidean distance between data point x and centroid cj.

The primary objective of the K-Means algorithm is to minimize the total squared distance between each data point and its cluster's centroid. A smaller value of J indicates better clustering quality and more compact clusters.

FINDINGS AND DISCUSSION

Table 2. Descriptive Statistics

Variable	Minimum	Maximum	Mean	Std. Deviation
MSMEs	5.84	977.47	132.39	227.09
SOSMED	3871.00	9808.00	8667.38	943.43
E-COM	1820.00	5679.00	3459.50	982.21
QRIS	1.18	932.00	274.47	260.06
LABOR	5695.00	8706.00	7380.61	714.52

Source: Data Processed

Based on the descriptive statistics in Table 2, the MSMEs variable has a minimum value of 5.84 and a maximum of 977.47 with an average of 132.39 and a standard deviation of 227.09, indicating a large variation between regions. The SOSMED variable has an average of 8,667.38 with a range of 3,871 to 9,808 and a standard deviation of 943.43, indicating a high and relatively even level of social media usage. E-commerce adoption (ECOM) recorded a mean of 3,459.50, a minimum of 1,820, a maximum of 5,679, and a standard deviation of 982.21, indicating a fairly wide spread. The use of QRIS shows an average value of 274.47 with a high variation (standard deviation 260.06) from a minimum of 1.18 to a maximum of 932.00, indicating inequality in utilization between regions. Meanwhile, the amount of labor (LABOR) is relatively stable with an average of 7,380.61, a minimum of 5,695, a maximum of 8,706, and a standard deviation of 714.52. This data provides an initial picture of the distribution and diversity of variables that will be analyzed further in the clustering process.

Table 3. Multicollinearity test

Model	Tolerance	VIF
Zscore Sosmed	0.681	1.469
Zscore E-com	0.643	1.554
Zscore Qris	0.946	1.057
Zscore Labor	0.820	1.219

Source: Data Processed, 2025

Based on the results of the statistical analysis presented in Table 3, observations show that all research variables have tolerance values that exceed the 0.1 threshold and Variance Inflation Factor (VIF) values that are below the critical value of 10. This finding indicates that there is no multicollinearity problem in the set of variables analyzed. In order to form an optimal cluster structure, a data standardization procedure was carried out to generate Z-score values that serve as the basis for clustering. The transformation of these values allows for an in-depth analysis of the characteristics and distribution patterns in each of the clusters formed, thus providing a comprehensive interpretation of the clustering structure of the data.

Table 4. Initial Cluster Center

	Cluster		
	1	2	3
MSMEs	3.721	-.230	-.523
SOSMED	-.253	1.209	-5.083
E-COM	.880	2.259	-1.324
QRIS	-1.042	-1.035	-.374
LABOR	-.443	1.450	1.043

Source: Data Processed, 2025

Table 4 displays the initialization results of the clustering process, which represents the initial stage of forming the clustering structure before the iteration procedure. The data documented in the table is the fundamental foundation for the formation of the three clusters targeted by the study. This stage serves as an algorithmic starting point that facilitates the grouping of research objects into three separate categories, before the iterative process of optimizing the position and characteristics of each cluster.

Table 5. Iteration History

Iteration	Change in Cluster Centers		
	1	**2**	**3**
1	2.243	3.091	.000
2	1.693	.120	.000
3	.000	.000	.000

Source: Data Processed, 2025

Table 5 illustrates the series of iterative processes in cluster formation, which includes three consecutive iteration cycles. In the first iteration, the analysis shows that the centroid change has not reached the required significance level. A similar phenomenon occurs in the second iteration, where the centroid shift still falls below the significance threshold. However, in the third iteration, the analysis showed that the centroid changes reached the required level of significance. This significant transformation in the centroid position in the third iteration indicates the achievement of stability in the clustering process and confirms the validity of the cluster structure formed.

Table 6. Final Cluster Center

	Cluster		
	1	**2**	**3**
MSMEs	3.060	-.288	-.523
SOSMED	-.078	.177	-5.083
E-COM	.839	-.039	-1.324
QRIS	-1.038	.116	-.374
LABOR	.955	-.130	1.043

Source: Data Processed, 2025

Table 6 presents the final results of the clustering analysis process that has been carried out, displaying the formation of three clusters for each research variable. The values listed in the final cluster centers table represent the results of the standardization of the processed data. The interpretation of these numbers can be explained through their polarity, where negative values indicate that the data is below the total population mean, while positive values indicate that the data is above the total mean. To identify the significance and magnitude of each variable's influence on cluster formation, further analysis was conducted by implementing the predetermined calculation formula.

$$X=\mu + z.\sigma$$

Description:

μ = population mean

σ = standard deviation

z = standardization value

Table 7. Cluster Membership

Number of cases	Province	Cluster	Distance
1	ACEH	2	.97242
2	NORTH SUMATRA	2	149.518
3	WEST SUMATRA	2	141.682
4	RIAU	2	186.203
5	JAMBI	2	.54858
6	SOUTH SUMATRA	2	211.304
7	BENGKULU	2	.78383
8	LAMPUNG	2	241.884
9	BANGKA BELITUNG ISLAND	2	113.917

Number of cases	Province	Cluster	Distance
10	RIAU ISLAND	2	202.726
11	DKI JAKARTA	2	319.084
12	WEST JAVA	1	.99617
13	CENTRAL JAVA	1	.93841
14	YOGYAKARTA	2	193.327
15	EAST JAVA	1	155.743
16	BANTEN	2	232.659
17	BALI	2	259.568
18	WEST NUSA TENGGARA	2	132.893
19	EAST NUSA TENGGARA	2	204.785
20	WEST KALIMANTAN	2	.27596
21	CENTRAL KALIMANTAN	2	129.919
22	SOUTH KALIMANTAN	2	.71011
23	EAST KALIMANTAN	2	128.906
24	NORTH KALIMANTAN	2	108.816
25	NORTH SULAWESI	2	169.941
26	CENTRAL SULAWESI	2	.62877
27	SOUTH SULAWESI	2	242.943
28	SOUTHEST SULAWESI	2	128.797
29	GORONTALO	2	195.437
30	WEST SULAWESI	2	166.992
31	MALUKU	2	215.243
32	NORTH MALUKU	2	143.767
33	WEST PAPUA	2	209.784
34	PAPUA	3	.00000

Source: Data Processed, 2025

In this study, a cluster analysis method was applied using data from 34 provinces in Indonesia. The analysis results in a non-hierarchical clustering that divides the provinces into three categories based on the level of adoption of MSMEs digitalization. The regions analyzed cover all provinces in Indonesia, from Aceh, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, and Lampung, to Bangka Belitung Islands and Riau Islands in the western part of Indonesia. Furthermore, the regions in Java included in this study include DKI Jakarta, West Java, Central Java, DI Yogyakarta, East Java, and Banten. Meanwhile, in eastern Indonesia, the study includes Bali, West Nusa Tenggara, and East Nusa Tenggara, which have different economic characteristics and MSMEs development from other regions.

In the Kalimantan region, the research involved West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, and North Kalimantan. On the island of Sulawesi, the provinces analyzed include North Sulawesi, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo, and West Sulawesi. Further to the eastern region, this study also includes provinces in the Maluku Islands, namely Maluku and North Maluku, as well as provinces in Papua Island, namely Papua and West Papua.

Figure 2. MSMEs digitization map resulting from cluster analysis
Source: Data Processed (2025)

To facilitate interpretation of the clustering results, a comprehensive visualization shows that the data is divided into three groups based on the distribution of MSMEs, the number of digitized workers, the level of social media usage, and the adoption of QRIS-based payment technology. Green provinces represent MSME digitization levels that are well above the national average, reflecting progress in digital technology integration that drives competitiveness and business ecosystem growth. Yellow regions indicate digitalization adoption that is still in a transitional stage and requires strengthening digital strategies, including social media optimization, QRIS system implementation, and HR capacity building. Meanwhile, the red region illustrates significant delays in the adoption of digital technology in the MSMEs sector, characterized by minimal utilization of social media, low implementation of QRIS, and limited workforce readiness, hindering the overall development of the MSMEs ecosystem.

Research by El et al (2024) shows that MSMEs in the Java Island region have great potential in the digitization process. This is supported by their active involvement in various business sectors, especially through the Community-Based Tourism (CBT) approach. The implementation of CBT allows MSME players to utilize digital technology to disseminate product and service information more widely, even to the international level, thus strengthening their visibility and competitiveness in the global market.

Maulana et al (2024) argued that the stagnation of digitalization of MSMEs in areas outside Java, such as Sumatra, Kalimantan, and Papua, is due to the centralization of policy direction and digital technology development in the Java region. This imbalance causes limited access to internet infrastructure and low labor capacity in operating digital technology, thus slowing down the digital transformation that should be evenly distributed nationally. The underdevelopment classification assigned to the Papua region, especially in Southwest Papua Province, reflects the structural gap between the advancement of digital technology and the digital capabilities of indigenous Papuan micro, small and medium enterprises (MSMEs). This imbalance can be seen from the low level of literacy and mastery of technology by business actors in adopting digital platforms, especially in the context of online-based marketing and market development strategies (Maryen et al., 2024).

In this case, the spread of MSMEs in digitalization must be done starting from providing literacy and in-depth knowledge to the community. If the human resources are qualified, then digital readiness generally measures the readiness of the supporting components of digital transformation which involves indicators of People in carrying out the digitization process, the readiness of the MSMEs Process to digitize and the readiness of Technology that has been used and owned by MSMEs actors to carry out the transformation process and implementation of digital services.

CONCLUSIONS

This research objectively aims to segment regions in Indonesia based on the level of digitalization of MSMEs using the K-Means Clustering approach on five main indicators: number of MSMEs, number of e-commerce players, workforce in the e-commerce sector, QRIS adoption rate, and social media usage. The analysis shows three regional clusters with different digitalization characteristics: advanced cluster (West Java, Central Java, East Java), developing cluster (30 other provinces), and lagging cluster (Papua). This finding indicates a significant geographical gap in the digital transformation of the MSMEs sector in Indonesia.

Practically, the results of this mapping provide a strategic basis for policymakers, especially central and local governments, in designing more targeted, region-based MSMEs development policies. Areas with low digitalization require interventions in the form of digital training, improved technological infrastructure, and expanded access to digital platforms and cashless payment systems. Meanwhile, regions that are already in the high digitalization cluster can be encouraged to become national-scale MSMEs digital incubation centers.

Conceptually, this study makes a significant contribution to the development of an analytical framework in understanding the application of the K-Means Clustering method as a quantitative approach to mapping the level of digitalization readiness of MSMEs in the regional dimension. The utilization of this method not only allows the identification and classification of regional characteristics based on digital indicators objectively, but also builds a methodological foundation that can support evidence-based public policy making. Therefore, this approach has strategic relevance in supporting sustainable digital economic transformation.

LIMITATION & FURTHER RESEARCH

This study has several limitations that need to be considered in interpreting the results and developing further studies. First, the scope of analysis is limited to 34 provinces in Indonesia without taking into account the latest administrative area developments. Second, this study uses a cross-sectional data approach that only represents the condition of MSMEs digitalization in 2023, so it cannot illustrate the dynamics or trends in digitalization changes over time. Third, the indicators used in the segmentation of MSMEs digitalization are still limited to five main variables, namely MSMEs distribution, number of e-commerce players, e-commerce workforce, QRIS usage, and social media. This study has not covered other important dimensions such as the level of digital literacy, managerial capacity, regional policies, or access to digital financing. Finally, this study does not directly evaluate the causal relationship between the level of digitalization and the economic performance or sustainability of MSMEs.

Future research is recommended to develop analysis using time series data in order to capture longitudinal patterns and trends of digitalization. The addition of contextual variables such as digital literacy index, technology perception, infrastructure readiness, and local policy intervention is also important to provide a more complete picture. In addition, integrative approaches with other methods such as geospatial analysis, fuzzy clustering, or qualitative case studies, can also be used to complement statistical mapping.

REFERENCES

Anatan, L., & Nur, T. (2023). E-commerce adoption and MSME performance: Evidence from Indonesia. *Journal of Small Business Management, 15*(2), 123–140.

ASPI. (2023). *Buletin Statistik ASPI Triwulan IV*. Asosiasi Sistem Pembayaran Indonesia.

Ayem, S., & Wahidah, U. (2021). Analisis faktor-faktor yang mempengaruhi kinerja keuangan UMKM di Kota Yogyakarta. *Jurnal Riset Akuntansi dan Keuangan, 9*(1), 45–62.

Ayu Nursasi, D., Pratiwi, S., & Rahman, A. (2024). Digitalization barriers and productivity challenges in Indonesian MSMEs. *Asian Journal of Business Research, 12*(3), 78–95.

Basri, H., Mulyadi, D., & Sari, P. (2023). Digital transformation impact on MSMEs growth in Indonesia. *Indonesian Economic Review, 8*(2), 156–171.

Bock, H. H. (2008). Origins and extensions of the k-means algorithm in cluster analysis. *Electronic Journal for History of Probability and Statistics, 4*(2), 1–18.

BPS. (2024). *Jumlah Perusahaan Industri Skala Mikro dan Kecil Menurut Provinsi (Unit)*. https://www.bps.go.id/id/statistics-table/2/NDQwIzI=/jumlah-perusahaan-industri-skala-mikro-dan-kecil-menurut-provinsi.html

Drummond, C., McGrath, H., & O'Toole, T. (2020). Digital marketing capabilities in SMEs: The role of social media utilization. *European Journal of Marketing, 54*(6), 1234–1256.

Eka Ratnawati, S., Kusumawati, A., & Pratama, B. (2014). Clustering analysis for regional development planning using K-means algorithm. *Journal of Regional Science, 6*(2), 89–105.

El Yamin, D. A., Noor, S. F. S. J., Salsabila, I., Khairani, J. N., Putri, N. A., & Pratiwi, N. A. (2024). Dampak pariwisata dalam mewujudkan kesejahteraan masyarakat di Desa Wisata Karangrejo melalui Community Based Tourism (CBT). *Jurnal Nasional Pariwisata, 14*(2), 108–125. https://doi.org/10.22146/jnp.100584

Gujarati, D. (2015). *Econometrics by example*. Bloomsbury Publishing.

Handini, S., & Choiriyati, W. (2019). Digitalisasi UMKM sebagai hasil inovasi komunikasi pemasaran di masa pandemi COVID-19. *Jurnal Komunikasi Bisnis, 8*(3), 201–215.

Idris, A. (2024). Opportunities and strategies for accelerating digital transformation: Spatial distribution analysis and clustering using K-medoids method in South Sulawesi, Indonesia. *Journal of Mathematics and Applied Statistics, 2*(1), 51–62. https://doi.org/10.35914/mathstat.v2i1.188

ILO, & OECD. (2022). *Digital transformation challenges for MSMEs in developing countries*. International Labour Organization.

Kumar, S., Singh, A., & Patel, R. (2020). Data mining techniques for business intelligence: A comprehensive approach. *International Journal of Data Science, 7*(4), 234–251.

Kurnia, A. A., & Wulandari, D. (2022). Perbandingan UMKM yang memanfaatkan digitalisasi dan non-digitalisasi di Lamongan pada era COVID-19. *Co-Creation: Jurnal Ilmiah Ekonomi Manajemen Akuntansi dan Bisnis, 1*(2), 80–94. https://doi.org/10.55904/cocreation.v1i2.306

Lv, C., Zhang, Y., & Wang, L. (2025). Human capital dimensions in digital transformation: Evidence from e-commerce sector employment. *Technology and Society Review, 18*(1), 45–62.

Maulana, I. R., Kalila Salma, N., & Surachman, A. R. R. (2024). Optimalisasi pengembangan start-up dalam mengatasi middle-income trap era Jokowi: Analisis kebijakan holistik dan sinergi multisektor. *Indonesia Foreign Policy Review, 11*(1), 54–78. https://doi.org/10.5281/zenodo.14562302

Maryen, A., Clan, E., Patiasina, R., & Manajemen, J. (2024). Pelatihan digital marketing pada UMKM Orang Asli Papua di Provinsi Papua Barat Daya. *Journal of Human and Education, 4*(1), 257–263.

Montesi, M., & Urdiciain, B. G. (2005). Classification techniques in information retrieval research. *Knowledge Organization, 32*(2), 84–91.

Pakpahan, A. S. (2021). Efisiensi dan efektivitas digitalisasi pencatatan keuangan UMKM. *Jurnal Akuntansi UMKM, 5*(2), 67–82.

Remawati, D., Putra, D. J. A., & Irawati, T. (2021). Metode K-means untuk pemetaan persebaran usaha mikro kecil dan menengah. *Jurnal Teknologi Informasi dan Komunikasi (TIKomSiN), 9*(2), 39–44.

Reski, M. (2024). Pentingnya penyusunan laporan keuangan UMKM dalam era digital. *Jurnal Keuangan dan Bisnis, 12*(1), 34–48.

Rogers, E. M. (2024). *Diffusion of innovations* (6th ed.). Free Press.

Rusdiaman Rauf, Syam, A., & Randy, M. F. (2024). Optimalisasi transformasi digital dalam mendorong pertumbuhan usaha mikro, kecil dan menengah di Indonesia. *Bongaya Journal for Research in Management, 7*(1), 95–102. https://doi.org/10.37888/bjrm.v7i1.594

Tazkia, R., Putri, S., & Hassan, M. (2024). QRIS adoption patterns among Indonesian MSMEs: A cluster analysis approach. *Digital Payment Research, 11*(3), 178–195.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science, 46*(2), 186–204.

Verma, M., Srivastava, M., Chack, N., Diswar, A. K., & Gupta, N. (2012). A comparative study of various clustering algorithms in data mining. *International Journal of Engineering Research and Applications, 2*(3), 1379–1384.

Wijayantini, S., Pratama, A., & Sari, D. (2024). Government policies and digital transformation acceleration in Indonesian MSMEs. *Public Policy Review, 15*(2), 89–104.

Yadohisa, H., Tanioka, K., & Fujiwara, T. (2010). Data standardization effects on clustering performance in multivariate analysis. *Statistical Methods and Applications, 19*(3), 245–260.