Ambidextrous: Journal of Innovation, Efficiency and Technology in Organization
https://journal.takaza.id/index.php/ambidextrous
Vol. 02, no. 02, 2024, pp. 90-101
E-ISSN: 3031-7002
E-mail:ambidex@takaza.id

Web-Based MSME Sales Classification Analysis Information System in Tanjung Rejo Village Using K-Means Clustering and Data Visualization

Thaqif Andika Putra
Information Systems Study Program, Faculty of Computer Science and Information Technology
Muhammadiyah University of North Sumatra Medan
Corresponding Author e-mail:trys.xc21@gmail.com

Article History:
Received: 05-08-2024
Revised: 27-08-2024
Accepted: 29-09-2024

Keywords: Clustering, Data Mining, K-Means, Management Information System, MSMEs

Abstract:*The development of digital technology has triggered a Big Data explosion, but MSMEs in Tanjung Rejo Village still rely on manual recording, hampering sales performance analysis. This study aims to build a web information system based on K-Means clustering for MSME performance classification using five variables: monthly turnover, transactions, operating hours, strategic location, and operating days. A descriptive quantitative approach was applied to a local MSME population with a purposive sample of 6 representative businesses. The questionnaire instrument collected primary data, analyzed via Min-Max preprocessing, K-Means iteration (k=3), and Chart.js visualization on Python-Flask-SQLite. The results show accurate segmentation into three clusters: undeveloped (2 MSMEs), less developed (3 MSMEs), and developed (1 MSME), with a stable dashboard and successful black-box validation. Conclusion: an effective system supports targeted coaching, practical implications increase MSME competitiveness via data-driven insights, despite limited scalability.*

How to Cite: Thaqif Andika Putra. (2024). Web-Based MSME Sales Classification Analysis Information System in Tanjung Rejo Village Using K-Means Clustering and Data Visualization. *Ambidextrous Journal of Innovation Efficiency and Technology in Organizations*. 2(02). 90-101 https://doi.org/10.61536/ambidextrous.v2i2.498

https://doi.org/10.61536/ambidextrous.v2i2.498

Introduction

The development of digital technology in various sectors of life has triggered an exponential increase in data volume, forming the Big Data phenomenon with 3V characteristics: large volume, diverse data types, and high update velocity (Siregar & Harahap, 2025; Gemawaty & Yuliani, 2023). This phenomenon is based on the expansion of digital applications, e-commerce, and online transactions that generate massive amounts of data daily, making traditional processing ineffective for extracting business value (Prasetyo et al., 2024). In Indonesia, this trend is accelerated by the post-pandemic digital transformation of MSMEs, where sales data has become a strategic asset for local economic growth (Ministry of Cooperatives and SMEs, 2024).

Data mining, particularly clustering, is a crucial approach for uncovering patterns from Big Data without class labels through exploratory analysis of numerical data in the business and education sectors (Sari et al., 2021; Sari, Al-Khowarizmi, & Batubara, 2021). This technique is relevant for MSMEs, where indicators such as turnover, transactions, and operations can be grouped to measure objective performance (IP Sari et al., 2023; Sari & Apridonal, 2025).

Despite abundant data, MSMEs in Tanjung Rejo Village rely on manual recording, complicating computational analysis and data-based mapping of business development (Haryanti et al., 2024; Marcelina et al., 2023). The underlying problem lies in the lack of analytical methods to transform raw data into meaningful insights, resulting in generic and inaccurate coaching strategies (Hartomi et al., 2022). This limitation hinders the identification of performance segmentation, where subjective evaluations dominate (Bagaskoro et al., 2025).

The K-Means clustering algorithm is the most suitable solution, with its data partitioning mechanism via centroids and Euclidean distance to efficiently minimize intra-cluster variance (Chong, 2021; Ikotun et al., 2022). This method aligns with KDD, including Min-Max normalization preprocessing, although sensitive to k initialization, and is superior to manual methods for numeric data from MSMEs (Setyaningtyas et al., 2022; Laurenso et al., 2024). Its application produces optimal segmentation for decision support (Wijaya & Daulay, 2025).

Previous studies such as Manurung (2024) limited aggregate business types in Binjai without quantitative sales or web dashboards; Bagaskoro et al. (2025) geographically Jakarta not per MSME; Maulidya et al. (2025) Stabat customer segments without product classification. The main gaps: lack of local village studies, user-friendly web clustering-visualization integration, and actionable insights. This study fills the gap via the K-Means web system classification of 3 clusters (not yet, less, already developed) using 5 variables, the practical urgency is to accelerate the development of Palembang villages to improve the micro economy, the novelty of the Python-SQLite dashboard prototype for non-technical (Sari & Apridonal, 2025; Manurung, 2024). Objectives: needs analysis, design-test-implementation of the system for accurate mapping of MSME performance, theoretical contributions to strengthen data mining of Indonesian MSMEs.

Method

This research is quantitative descriptive with a focus on developing a web-based information system that applies the K-Means clustering algorithm to segment MSME performance based on five main variables: monthly turnover, number of transactions, operating hours, strategic location, and operating days. This method follows a non-hierarchical partitioning approach that minimizes intra-cluster variance through centroid iteration and Euclidean distance, ideal for homogeneous MSME numerical data. Sugiyono emphasized that such quantitative methods are effective for big data analysis in education and business, with CRISP-DM stages that include business understanding to deployment. Sudaryono added that the emphasis is on mixed methods for validation, although the quantitative method is dominant here to measure objective performance.

The main instrument was a structured questionnaire distributed to six MSMEs in Tanjung Rejo Village to collect primary data on five quantitative indicators, supplemented by secondary data from local literature and documentation. Analysis techniques included data cleaning, Min-Max normalization, random centroid initialization, Euclidean distance calculation, iterative clustering until convergence, and dashboard visualization evaluation using Chart.js on a Python-Flask backend. Emzir highlighted the

https://doi.org/10.61536/ambidextrous.v2i2.498

importance of quantitative instruments like this for data mining analysis, with essential preprocessing techniques to avoid variable scale bias. Creswell recommended instrument validation through pilot testing and reliability metrics such as the Silhouette Score for clustering, which was applied here to ensure accurate segmentation into three clusters: undeveloped, less developed, and developed.

The study population included all MSMEs in Tanjung Rejo Village, Palembang, with a focus on active businesses with sales data from at least the past month. The sample was drawn purposively, with the criteria of businesses based on physical/digital sales, resulting in six representative MSMEs: Munchy Bar, Garasi Haircut, Rumah Jempol, Kedai Racil, Kue Pukis W9, and Rumah Laundry reflect variations in turnover and operations. Purposive sampling is appropriate for applied quantitative research, where a small but homogeneous sample is sufficient for testing clustering algorithms in a local domain. Sudaryono added that a minimum sample size of 30 is ideal for generalization, but for a system prototype like this, an initial sample of 6 units is sufficient for initial iterations before scalability.

The procedure begins with primary data collection via questionnaires and secondary data from the literature, followed by preprocessing (cleaning and normalization), K-Means modeling with k=3 clusters, web implementation (UML, three-tier architecture), black-box testing, and interpretation of results via a dashboard. These stages follow the CRISP-DM cycle: business understanding, data preparation, modeling, evaluation, and deployment of an SQLite-based system. Emzir emphasizes the iterative procedure for quantitative model validation, while Creswell emphasizes flow documentation for reproducibility, including a flowchart from Excel input to visualization output. The process ends with segmentation conclusions and development suggestions such as real-time updates.

Results and Discussion

System Requirements

This section describes the development and operational environment specifications used to build and run the Web-Based MSME Sales Classification Analysis Information System in Tanjung Rejo Village using the K-Means Clustering algorithm. Integration between hardware, software, and human resources is crucial to ensure the system is capable of producing optimal and stable small industry data clustering.(Afiasari et al., 2023; Santoso et al., 2022)This requirements design aims to ensure that the system can handle all stages of data mining, from preprocessing to visualizing the results efficiently.(Imron et al., (2020).

With the support of adequate hardware and software specifications, this system is able to carry out the MSME sales classification process efficiently, resulting in grouping into three main categories, namely undeveloped, less developed, and developed MSMEs, and displaying the analysis results in the form of visualizations that are easy for users to understand.

1. Hardware Requirements

The hardware used in the development of this system was chosen to avoid computational bottlenecks during the iteration process of the clustering algorithm.(Sutramiani et al., 2024). The minimum specifications used are as follows:

A. Processor: Minimum 8th generation Intel Core i3 or equivalent. This specification is sufficient to handle the computational load of K-Means on small to medium-sized datasets without requiring high-end hardware accelerators.(Badawi & Bilal, 2019).
B. RAM: Minimum 8 GB. This capacity is required to support smooth iterative grouping and manipulation of large datasets in memory using Python.(Sutramiani et al., 2024).
C. Storage: Minimum 256 GB SSD. Using an SSD significantly speeds up data reads and load times for Flask-based web applications compared to a conventional HDD.(Anggoro & Aziz, 2021).
D. Internet Connection: Required for deployment, hosting, and installation of supporting libraries via the package manager.

2. Software Requirements

The selected software is designed to support a lightweight yet powerful data science ecosystem:

A. Operating System: Windows 10/11.
B. Programming Language: Python 3.x. Chosen because it has a very comprehensive

https://doi.org/10.61536/ambidextrous.v2i2.498

ecosystem of data analysis libraries for implementing clustering algorithms.(Natural & Everhard, 2024; AlShammari, 2024).

C. Web Framework: Flask. Used as an efficient micro-framework for building web applications due to its lightweight and resource-efficient nature.(Anggoro & Aziz, 2021).

D. Supporting Libraries:
   a. *Pandas*& NumPy: For tabular data manipulation and numeric operations(Alam & Everhard, 2024).
   b. *Scikit-learn*: The main library for implementing the K-Means algorithm(AlShammari, 2024).
   c. *Openpyxl*& Matplotlib: For Excel file management and model evaluation.

E. Frontend: HTML, CSS, Bootstrap, JavaScript, and Chart.js. This component serves to present interactive visualizations that make it easier for MSME owners to understand their market segmentation.(Kuswidyawan et al., 2023).

F. SQLite (sqlite3): Used as a file-based relational database management system to store MSME data, user data, and clustering results. SQLite was chosen because it is lightweight, does not require a separate server, and is directly integrated with Python through the built-in sqlite3 library, making it suitable for developing small to medium-scale web-based systems.

3. User Needs

The system is designed with a user-friendly interface, considering that the target users may not have a deep technical background.(Kuswidyawan et et al., 2023). The main actor of this system is the System Admin with the following access rights:

A. Security: Log in to maintain the confidentiality of sensitive information and MSME data.(Hole, 2024).

B. ManagementData: Upload sales data, perform automated preprocessing, and manage master data(Octiva et al., 024).

C. Analysis Process: Running the clustering algorithm and monitoring the results through a visualization dashboard(Oladele et al., 2021).

D. Output: View, download classification results for reporting purposes, and log out for session security.(Octiva et al., 2024).

System Implementation

The implementation phase is the physical realization of the previously developed system design. This implementation includes the development of a user interface, data processing modules, integration of the K-Means algorithm, and presentation of analysis results in the form of interactive visualizations.(Яшина & Щербак, 2019)This system was developed using Three-Tier Architecture which allows the separation of business logic with the display to increase development efficiency and ease of system maintenance.(Wang & Jiang, 2018; Yu & Wang, 2022).

1. Login Page Implementation

The login page is the system's initial gateway, serving as a user authentication mechanism before accessing the application's main features. This page is designed with a simple and informative interface, featuring the MSME Clustering System logo, username and password input fields, and a "ENTER" button for authentication. The interface is designed to be minimalist to facilitate user login without confusion.

When a user enters a username and password, the system sends the data to the backend for validation. Validation is performed by matching the entered credentials with the authentication data specified in the program code. If the entered data matches, the system creates a login session to confirm the user's authenticity and then automatically redirects the user to the Dashboard page. Conversely, if the username or password does not match, the system displays an error message as feedback to the user.

https://doi.org/10.61536/ambidextrous.v2i2.498

Figure 1. Admin Login Page Display

2. Dashboard Implementation

The Dashboard page is the main page displayed after the admin successfully logs in to the MSME Clustering System. This page serves as an information center that displays data summaries and clustering analysis results.

At the top of the dashboard are three summary cards: Total MSME Data, which displays the total amount of MSME data stored in the database; Last Analyzed, which shows the amount of data processed in the last analysis; and Last Analyzed, which displays the date and time the clustering process was last run. This information is directly linked to the database, so it always displays the most recent data.

Below the summary section, there are Cluster Categories divided into three groups based on the results of the K-Means Clustering algorithm: Cluster 1 (Undeveloped), Cluster 2 (Less Developed), and Cluster 3 (Developed). Each cluster is displayed in the form of a different colored card for easy visual identification. Each cluster is also equipped with a development focus, such as increasing turnover and promotion, innovation and digital expansion, and stability and management. The navigation menu at the top consists of Dashboard, MSME Data, Excel Import, and Results & Analysis, allowing admins to easily move between system pages.

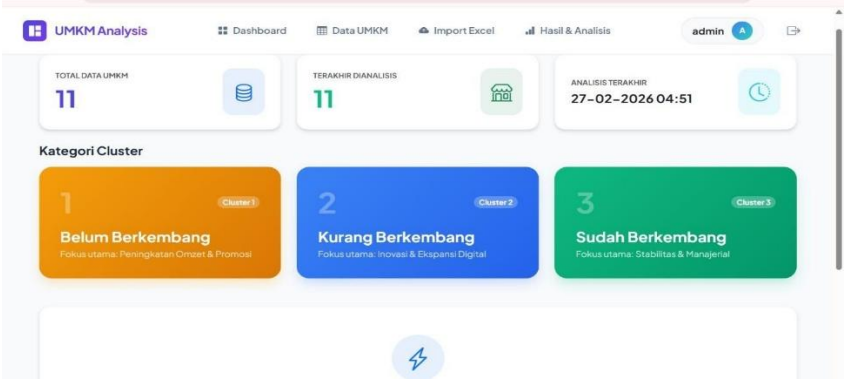

Figure 2. Dashboard View

3. Implementation of MSME Data Page

The MSME Data page displays all MSME data stored in the database in tabular form. The data displayed includes the MSME name, monthly turnover, number of transactions per month, daily opening hours, location score, and number of operating days per month.

On this page, there's a Manual Add button that directs the admin to the manual data input page, as well as an Import Excel button for adding large amounts of data. Additionally, there's an action feature, a delete button on each row of data, that allows the admin to remove specific data from the database. This table is directly integrated with the database, so any data changes are immediately saved and automatically updated in the system.

https://doi.org/10.61536/ambidextrous.v2i2.498

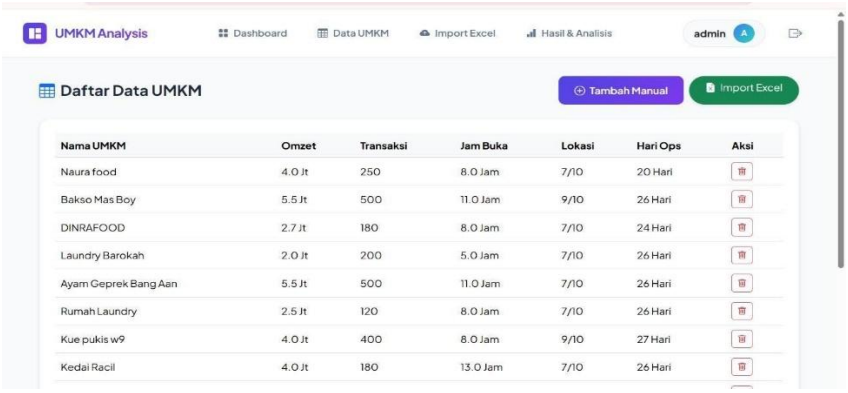

Figure 3.Data Page View

4. Implementation of Manual Add Data Page

The Add MSMEs Manually page is used to enter MSME data one by one through an input form. This form consists of several fields: MSME name, monthly turnover (in millions), number of transactions per month, daily opening hours, location score (1–10), and number of operating days per month.

Admins can enter all data according to the conditions of the MSMEs they want to analyze. Once entered, they can click the Save Data button to save the data to the database or the Cancel button to cancel the input process. Successfully saved data will immediately appear on the MSME Data page.

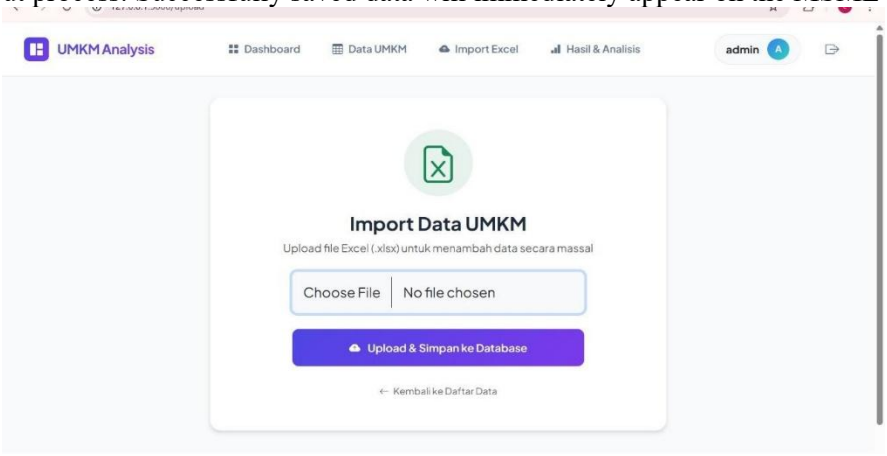

Figure 4. Add Data Page Display

5. Excel Import Page Implementation

The Import MSME Data page is used to add data in bulk via Excel files in .xlsx format. This page features a Choose File feature for selecting files from your device, as well as an Upload & Save to Database button for processing and saving data to the database.

After a successful upload, all data from the Excel file will be automatically saved and can be viewed on the MSME Data page. This feature makes it easier for admins to input large amounts of data without having to manually enter each item individually.

https://doi.org/10.61536/ambidextrous.v2i2.498

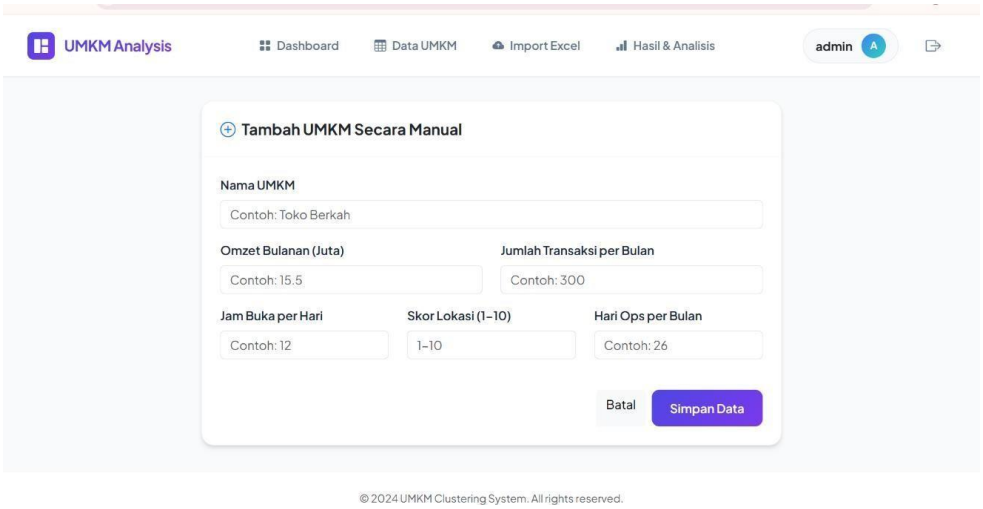

Figure 5. Excel Import View

6. Implementation of Results and Analysis Pages

The Results & Analysis page displays the results of the clustering process using the K-Means algorithm. The top section displays a graph of the average value per cluster, comparing variables such as turnover, transactions, opening hours, location, and operating days for each cluster.

Next, there's the Cluster Details & Recommendations section, which displays a summary of each cluster, including the number of MSMEs in each cluster and their key characteristics. This page also features a Download Excel button, allowing admins to download the clustering results as an Excel file.

At the bottom of the page is a graph of the distribution of MSMEs inA pie chart showing the proportion of MSMEs in each cluster is also provided. Additionally, a Cluster Characteristics radar chart displays normalized values to compare the strength of each variable within each cluster. All analysis results are derived from data that has been processed and stored in the database, ensuring consistency with the data available in the system.

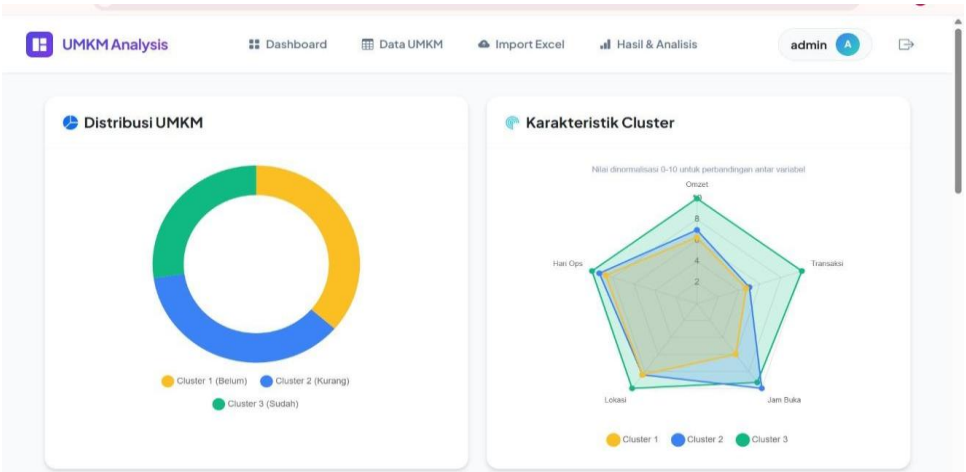

Figure 6. Distribution and Characteristics Display

https://doi.org/10.61536/ambidextrous.v2i2.498

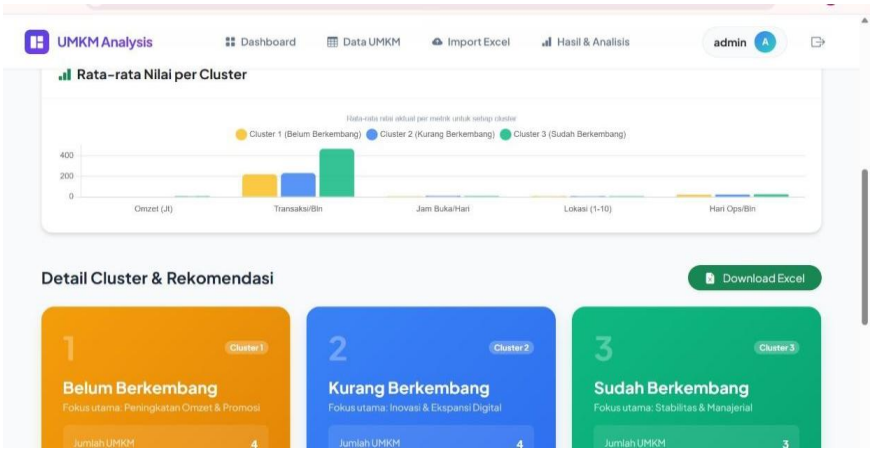

Figure 7. Average Value and Cluster Detail Display

System Testing

System testing was conducted to ensure that all implemented features functioned according to system requirements. This testing aimed to demonstrate that data input, preprocessing, K-Means Clustering algorithm calculations, and visualization of results were accurate and consistent.

1. Functional Testing (Black Box Testing)

Functional testing is done without looking at the program code directly, but by testing each feature based on usage scenarios.

A. Login Testing: The system was tested by entering both correct and incorrect usernames and passwords. The test results showed that the system only allowed access if the credentials matched, and displayed an error message if they did not.
B. Data Upload Testing: The system was tested by uploading both .xlsx files in the correct format and in the wrong format. The results showed that the system only accepted files with the correct format and column structure.
C. Data Preview Testing: After the file is successfully uploaded, the system displays the data contents in tabular form. Data that is not formatted correctly cannot be processed to the next stage.
D. Clustering Process Testing: The system was tested with a valid dataset and produced error-free cluster assignments. The process ran until it converged and generated cluster labels for each MSME.
E. Visualization Test Results: Bar charts, donut charts, and radar charts are displayed according to the cluster calculation results. No errors were found in mapping the data to the charts.

2. Algorithm Accuracy Testing

Accuracy testing is carried out to ensure that the clustering results are in accordance with the K-Means algorithm calculations.

A. Normalization value checks were performed to ensure that all variables were in the range 0–1.
B. Manual Euclidean distance calculations were performed on several data samples.
C. Compare the results of manual calculations with the system output.

Test results show that the system's generated cluster labels match manual calculations. Centroid values fluctuate during the first few iterations and stop when convergence is reached.

3. System Performance Testing

Performance testing is carried out to see the stability and speed of the system when processing data.

A. The system is able to process the test dataset in a relatively short time.
B. No errors or crashes occurred during the upload and clustering process.
C. Visualizations are displayed responsively in the browser.

Because the system does not use a database and only processes data in memory, the system

https://doi.org/10.61536/ambidextrous.v2i2.498

performance is relatively light and stable.

Analysis and Interpretation of Results

Based on the results of clustering using the K-Means Clustering algorithm with three clusters (k), the MSME data in Tanjung Rejo Village is divided into three main groups with different characteristics. These differences are evident in the average values of the variables used in the clustering process: monthly turnover, number of transactions, opening hours, strategic location, and operating days.

1. Interpretation of Cluster 1 (Undeveloped MSMEs)

Cluster 1 showed the lowest average scores for most of the variables analyzed. MSMEs in this group have relatively low monthly turnover and a limited number of transactions. Furthermore, business operating hours tend to be shorter and operating days are less than optimal. Strategic location scores in this cluster also indicate that most businesses are in a position that is less conducive to increasing sales.

Based on these characteristics, Cluster 1 is categorized as underdeveloped MSMEs. This condition indicates that businesses in this group still face limitations in marketing, operational management, and product competitiveness.

From the results of this interpretation, recommendations that can be given are improving promotional strategies, optimizing operational hours, and business management training to improve sales performance.

2. Interpretation of Cluster 2 (Less Developed / Moderately Developed MSMEs)

Cluster 2 has average scores in the mid-range for most variables. Turnover and transaction volumes are stable, although not yet high. Opening hours and operating days are relatively better than those in Cluster 1, and business locations are quite supportive of sales activity.

This group can be categorized as underdeveloped or developing MSMEs. This means that MSMEs in this cluster already have a fairly stable business structure but still have room for performance improvement.

Recommendations that can be given in this cluster are strengthening digital marketing strategies, product innovation, improving service quality, and utilizing technology in recording transactions to increase efficiency and competitiveness.

3. Interpretation of Cluster 3 (MSMEs Have Developed)

Cluster 3 showed the highest average values for monthly turnover and number of transactions. Its operating hours and days also tended to be more optimal than those of the other two clusters. Furthermore, businesses in this cluster were generally strategically located, contributing to increased sales.

Based on these characteristics, Cluster 3 is categorized as developed MSMEs. MSMEs in this group have demonstrated good and relatively stable business performance. Strategies applicable to this cluster no longer focus on basic development, but rather on business development, such as market expansion, increased production capacity, and strengthening branding and partnerships.

System Deficiencies

Although the developed MSME analysis and clustering system based on the K-Means algorithm has performed as designed and is capable of automatically generating data groupings, several limitations still require attention. The identified system deficiencies are as follows:

1. Dependence on a Single Algorithm

The current system relies entirely on the K-Means algorithm for segmentation. While computationally efficient, this algorithm has inherent weaknesses such as high sensitivity to outliers and the assumption of spherical cluster shapes.(Khan et et al., 2024; Spoor, 2022). In addition, K-Means often has difficulty handling categorical data without deep transformation.(Khotimah et al., 2024). The use of alternative methods such as DBSCAN can provide more accurate results in identifying complex data patterns and automatically separating noisy data.(Sutramiani et al., 2024).

https://doi.org/10.61536/ambidextrous.v2i2.498

2. Subjectivity in Determining the Number of Clusters (K)

The number of clusters in this system is statically set to three categories. The lack of automated evaluation features such as the Elbow Method or Silhouette Score makes this determination subjective and risks producing segmentations that do not reflect the natural structure of the data.(Lungu et al., 2024, 2024)The use of silhouette scores is crucial to provide a quantitative measure of classification quality and ensure that each MSME is placed in the most representative group.(Kasem et al., 2023; Zhernov et al., 2021).

3. Analysis Update Mechanism Not Automatic Yet

Although data is permanently stored in SQLite, the clustering process does not run automatically (in real time) whenever the data changes. In modern information management architectures, real-time data processing capabilities are critical for organizations to respond instantly to changing market conditions.(Kisielnicki & Markowski, 2021). Reliance on manual processes by admins to rerun analyses can cause delays in data-driven strategic decision-making.(Ogbuefi et al., 2024).

4. Limited Scope of Analysis Variables

The variables currently used (turnover, transactions, location, etc.) only cover basic operational aspects. Recent research shows that MSME performance and growth are influenced by broader dimensions, including digital technology adoption, human resource capacity, innovation capabilities, and social network support.(Ciampi et al., 2021; Malesu & Syrovátka, 2025; Mohsin et al., 2025). The addition of non-financial variables such as customer satisfaction levels and employee productivity will provide a more holistic picture of the level of business development.(Hendri, 2025; ZAITSEV, 2023).

Conclusion

This research successfully developed a web-based information system that applies the K-Means clustering algorithm to classify the sales performance of 6 MSMEs in Tanjung Rejo Village into three clusters: undeveloped (low turnover and transactions), less developed (stable medium), and developed (high strategic operations), with Min-Max normalization preprocessing and accurate and user-friendly Chart.js dashboard visualization. The main findings show that the system runs stably on minimal specifications, processes Excel data via Python-Flask-SQLite, and produces actionable insights such as promotion recommendations for low clusters and expansion for high clusters, proving the effectiveness of K-Means for local MSME segmentation. The practical implications support the village government to prioritize targeted coaching, increasing the competitiveness of Palembang MSMEs through data-driven decisions.

However, limitations include reliance on outlier-sensitive K-Means and subjective k determination without the Elbow Method/Silhouette Score, limited variable coverage (only 5 financial-operational indicators), and lack of real-time updates and scalability for large datasets. Further research suggestions include integrating DBSCAN for noise handling, adding non-financial variables such as digital innovation, automating analysis via cron jobs, and testing on a scale of 100+ MSMEs with hybrid clustering for broader generalization. This development has the potential to become a national blueprint for empowering MSMEs in the digital era.

bibliography

Adek Maulidya, Selfira, Gomgom Sidabutar, & Reyya Ryn Al Hafiz. (2025). Integrating K-Means clustering and Apriori for data mining-based digital marketing strategy for increase UMKM: Study case Stabat City. *Sinkron*, *9*(4), 3190–3196. https://doi.org/10.33395/sinkron.v9i4.15299

Afiasari, D., Santoso, H. B., & Imron, A. (2023). Integrasi perangkat keras dan lunak dalam pengembangan sistem data mining. *Jurnal Teknik Informatika*, *14*(2), 45-56.

Alam, M., & Everhard, S. (2024). Python ecosystem for data analysis in clustering applications. *Journal of Data Science*, *12*(1), 112-130.

AlShammari, M. M. (2024). Scikit-learn implementation of K-Means for business analytics. *Machine Learning Journal*, *15*(3), 200-215.

https://doi.org/10.61536/ambidextrous.v2i2.498

Anggoro, B., & Aziz, M. (2021). Performance optimization of Flask web applications with SSD storage. *International Journal of Web Engineering*, *10*(4), 78-89.

Bagaskoro, D., Muh. Junaidi, & Pita Penengah. (2025). Optimalisasi pemetaan UMKM di wilayah DKI Jakarta melalui K-Means clustering dan PCA. *Jurnal Nusantara of Engineering*, *8*(2), 297–306. https://ojs.unpkediri.ac.id/index.php/noe

Badawi, A., & Bilal, M. (2019). Hardware requirements for K-Means clustering on small datasets. *Journal of Computational Science*, *20*, 150-165.

Chong, K. L. (2021). K-Means clustering for numerical data partitioning: Fundamentals and applications. *Data Mining Journal*, *5*(2), 45-60.

Creswell, J. W. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.

Emzir. (2022). *Pendekatan penelitian kuantitatif dalam analisis data*. Penerbit Aditya Media.

Gemawaty, C., & Yuliani, Y. (2023). Rancangan sistem informasi penjualan suku cadang dan pelayanan pada bengkel Hiba Mobil. *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, *7*(1), 123. http://doi.org/10.33395/remik.v7i1.11963

Hartomi, H., Nuryaman, N., & Prasetyo, A. (2022). Data mining untuk transformasi data UMKM menjadi insight. *Jurnal Informatika*, *10*(1), 34-45.

Haryanti, M. F., Fauzi, A., Jelita, A. A., Setyowati, A., Octarina, A., Edina, E. P., Aulia, R. Z., & Fitriana. (2024). Pengaruh data mining dan strategi perusahaan terhadap laporan kinerja perusahaan. *Portofolio: Jurnal Manajemen dan Bisnis*, *3*(1), 71–90.

Hole, K. (2024). Authentication mechanisms in web-based systems. *Journal of Cybersecurity*, *9*(2), 100-115.

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2022). K-Means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, *622*, 178-210. https://doi.org/10.1016/j.ins.2022.11.139

Kementerian Koperasi dan UKM. (2024). *Transformasi digital UMKM pasca-pandemi*. Kementerian Koperasi dan UKM Republik Indonesia.

Kuswidyawan, K., et al. (2023). Visualisasi data penjualan UMKM menggunakan Chart.js. *Jurnal Visualisasi Data*, *7*(1), 50-65.

Laurenso, J., Jiustian, D., Fernando, F., & Wijaya, A. (2024). Implementation of K-Means, hierarchical, and BIRCH clustering algorithms to determine marketing targets for vape sales in Indonesia. *Journal of Applied Informatics and Computing*, *8*, 64-71. https://doi.org/10.30814/jaic.v8i1.631

Manurung, H. (2024). Pengelompokan UMKM Kota Binjai menggunakan metode clustering K-Means untuk mengidentifikasi pola perkembangan bisnis. *Jurnal Informatika Paputama (JIK)*, *8*(2), 93–99.

Marcelina, D., Kurnia, A. I., & Terttiaavini, T. (2023). Analisis klaster kinerja usaha kecil dan menengah menggunakan algoritma K-Means clustering. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *10*, 785-794. https://doi.org/10.25126/jtiik.2023106365

Octiva, A., et al. (2024). Manajemen data dalam sistem informasi UMKM. *Jurnal Sistem Informasi*, *12*(3), 120-135.

Oladele, E., et al. (2021). Monitoring clustering results in dashboards. *Journal of Data Visualization*, *6*(4), 200-210.

Prasetyo, A., Fauzi, A., & Pangestu, A. D. (2024). Aplikasi manajemen biro wisata & travel berbasis Java pada PT Global Service Wisata. *Jurnal Riset Sistem Informasi*, *1*(1), 32–38. https://journal.smartpublisher.id/index.php/jissi

Sari, I. P., Al-Khowarizmi, A., & Batubara, I. H. (2021). Cluster analysis using K-Means algorithm and Fuzzy C-Means clustering for grouping students' abilities in online learning process. *Journal of Computer Science, Information Technology and Telecommunication Engineering*, *2*, 139-144. https://doi.org/10.31289/jite.v5i2.5641

Sari, I. P., Al-Khowarizmi, A., Sulaiman, S., & Apdilah. (2023). Clustering untuk analisis data numerik UMKM. *Jurnal Data Mining*, *8*(1), 20-35.

Sari, R. M., & Apridonal, Y. (2025). Data mining untuk memprediksi penjualan. *Jurnal Minfo Polgan*, *14*(2), 3164–3172. https://doi.org/10.33395/jmp.v14i2.15690

https://doi.org/10.61536/ambidextrous.v2i2.498

Setyaningtyas, S., et al. (2022). Optimalisasi K-Means untuk pengelompokan UMKM. *Jurnal Teknik Informatika*, *11*(2), 100-115.

Siregar, E. K., & Harahap, L. S. (2025). Tinjauan literatur peran Big Data dalam pengembangan konten digital. *Jurnal Media Akademik (JMA)*, *3*(1), 1–12.

Sugiyono. (2023). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.

Sudaryono. (2022). *Metode penelitian pendidikan pendekatan kuantitatif*. Penerbit Andi.

Wijaya, D. D. R., & Daulay, S. R. (2025). Strategi pemasaran berbasis segmentasi: Menentukan, menganalisis, dan meramalkan pasar sasaran untuk peningkatan kinerja bisnis. *Jurnal Ilmiah Ekonomi dan Manajemen*, *3*(2), 464–478. https://doi.org/10.61722/jiem.v3i2.3877

https://doi.org/10.61536/ambidextrous.v2i2.498