Application of K-Means Clustering for Mapping and Analyzing Small and Medium Industry Potential Based on Production, Investment, and Labor Characteristics

M. Azwan, Muhammad Syahputra Novelan

Abstract

Small and Medium Industries (SMIs) play an important role in supporting regional economic growth, creating employment opportunities, and strengthening local economic development. However, differences in investment value, workforce size, and production performance create challenges in identifying and mapping the potential of each industrial unit. Therefore, a systematic analytical approach is required to group industries based on their characteristics and provide useful information for regional development planning. This study aims to apply the K-Means Clustering algorithm to map and analyze the potential of Small and Medium Industries based on production, investment, and labor characteristics. The research uses secondary data from the SMI database in Tanjungbalai City, North Sumatra, Indonesia, covering the 2021–2026 period. The initial dataset consists of 709 industrial records, with 708 complete records selected for the clustering process. Three numerical attributes are used as the main clustering variables, namely workforce size, investment value, and production value. The research stages include data collection, preprocessing, feature selection, data transformation, standardization, K-Means implementation, and cluster evaluation. The clustering results produce three groups: 502 records (70.90%) classified as low potential, 195 records (27.54%) as medium potential, and 11 records (1.55%) as high potential. Model evaluation generates a Davies-Bouldin Index (DBI) of 0.814 and a Silhouette Coefficient of 0.436. These findings indicate that K-Means Clustering can identify differences in industrial characteristics and provide structured information regarding the distribution of SMI potential. The results are expected to support government institutions in determining development priorities, improving investment capacity, strengthening workforce utilization, and formulating more targeted industrial development strategies based on data-driven decision-making.

***Keywords**: K-Means Clustering, Data Mining, Machine Learning, Small and Medium Industries, Industrial Potential.*

[1]M. Azwan
Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: azwanatlantica48@gmail.com

[2]Muhammad Syahputra Novelan
e-mail: putranovelan@dosen.pancabudi.ac.id
2nd International Cofference on the Epicentrum of Economic Global Framework (ICEEGLOF)
Theme: Navigating The Future: Business and Social Paradigms in a Transformative Era.
https://proceeding.pancabudi.ac.id/index.php/ICEEGLOF

Introduction

Industri Kecil dan Menengah (IKM) merupakan salah satu sektor penting dalam mendukung pertumbuhan ekonomi daerah, menciptakan lapangan pekerjaan, serta meningkatkan kesejahteraan masyarakat. Keberadaan IKM memiliki peran strategis dalam mengembangkan potensi ekonomi lokal melalui berbagai jenis kegiatan produksi, seperti industri pangan, sandang, logam, kerajinan, dan mebel. Perkembangan sektor IKM tidak hanya dipengaruhi oleh jumlah unit usaha, tetapi juga oleh kemampuan produksi, nilai investasi, serta jumlah tenaga kerja yang terlibat dalam kegiatan operasionalnya. Oleh karena itu, diperlukan pengelolaan data IKM yang baik agar potensi dan karakteristik setiap industri dapat diketahui secara lebih terstruktur sebagai dasar dalam menentukan arah pengembangan industri [1] [2].

Dalam praktiknya, pengelolaan data Industri Kecil dan Menengah masih menghadapi tantangan dalam mengidentifikasi tingkat potensi dan karakteristik masing-masing unit usaha. Data industri yang mencakup jenis produksi, cabang industri, jumlah tenaga kerja, nilai investasi, kapasitas produksi, dan nilai produksi memiliki karakteristik yang beragam sehingga sulit dianalisis hanya melalui pengamatan secara konvensional. Perbedaan nilai investasi, kapasitas produksi, serta jumlah tenaga kerja menyebabkan setiap IKM memiliki tingkat kemampuan dan potensi usaha yang berbeda. Apabila data tersebut hanya digunakan sebagai informasi administratif, maka informasi yang dihasilkan belum sepenuhnya mampu memberikan gambaran mengenai kelompok industri yang memiliki potensi tinggi, sedang, maupun rendah. Kondisi ini menunjukkan perlunya pendekatan analisis berbasis data untuk memperoleh informasi yang lebih terukur dan sistematis[3] [4] .

Pemanfaatan teknologi Data Mining dan Machine Learning dapat menjadi salah satu pendekatan untuk mengatasi permasalahan tersebut. Salah satu metode yang dapat digunakan adalah algoritma K-Means Clustering, yaitu metode pengelompokan data berdasarkan tingkat kemiripan karakteristik antarobjek ke dalam sejumlah kelompok atau cluster. Algoritma K-Means bekerja dengan menentukan sejumlah centroid sebagai pusat kelompok, kemudian menghitung jarak setiap data terhadap centroid untuk menentukan keanggotaan cluster. Dalam penelitian ini, algoritma K-Means dimanfaatkan untuk mengelompokkan IKM berdasarkan karakteristik produksi, nilai investasi, dan jumlah tenaga kerja. Melalui proses clustering, data industri yang memiliki karakteristik serupa dapat dikelompokkan sehingga menghasilkan informasi mengenai pola dan tingkat potensi masing-masing kelompok industri [5] [6] .

Penelitian ini menggunakan dataset Industri Kecil dan Menengah periode 2021–2026 yang diperoleh dari data direktori perusahaan IKM pada seluruh kecamatan di Kota Tanjungbalai. Berdasarkan dataset yang digunakan, terdapat 709 data industri dengan sejumlah atribut, antara lain jenis produksi, cabang industri, tenaga kerja, nilai investasi, kapasitas produksi, nilai produksi, dan tahun data. Variabel-variabel tersebut menjadi dasar dalam proses analisis dan pengelompokan potensi IKM. Tahapan penelitian meliputi pengumpulan data, preprocessing, seleksi atribut, normalisasi data, penentuan jumlah cluster, penerapan algoritma K-Means, serta interpretasi hasil pengelompokan. Hasil clustering selanjutnya dianalisis untuk mengetahui karakteristik setiap kelompok berdasarkan indikator yang digunakan, sehingga dapat memberikan gambaran mengenai distribusi potensi industri secara lebih objektif [7] [8].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan menerapkan algoritma K-Means Clustering untuk melakukan pemetaan dan analisis potensi Industri Kecil dan Menengah berdasarkan karakteristik produksi, investasi, dan tenaga kerja. Hasil penelitian diharapkan dapat menghasilkan pengelompokan IKM ke dalam beberapa kategori potensi berdasarkan karakteristik masing-masing unit usaha. Selain memberikan informasi mengenai pola pengelompokan industri, penelitian ini juga diharapkan dapat menjadi salah satu sumber informasi pendukung bagi pemerintah daerah dalam memahami kondisi sektor IKM dan menentukan prioritas pembinaan serta pengembangan industri. Dengan demikian, penerapan K-Means Clustering diharapkan mampu memberikan kontribusi dalam pemanfaatan teknologi

Machine Learning untuk mendukung analisis data industri dan pengambilan keputusan berbasis data[9].

Literature Review

1. Small and Medium Industries (SMIs)

Small and Medium Industries (SMIs) are important sectors in supporting regional economic development through production activities, employment opportunities, and the utilization of local economic resources. The development of SMIs can be assessed through several indicators, including the number of workers, investment value, production capacity, and production value. These indicators provide information about the operational characteristics and economic potential of each industrial unit. Differences in investment, workforce size, and production capacity indicate that each industry has different levels of business development and productivity. Therefore, analyzing these indicators is important to identify the characteristics and potential of SMIs systematically. In this study, SMI data are analyzed to identify groups of industries with similar production, investment, and workforce characteristics [10] [11].

2. K-Means Clustering Algorithm

Data Mining is a process of extracting useful information and identifying patterns from large datasets through statistical, computational, and machine learning techniques. One of the important techniques in Data Mining is clustering, which aims to group data objects based on similarities in their characteristics without requiring predefined class labels. Clustering can be applied to various fields, including business analysis, industrial development, economic mapping, and decision support systems. In the context of SMIs, clustering enables industrial units to be grouped according to similarities in investment value, production capacity, production value, and workforce size. Previous research by Marcelina et al. (2023) demonstrated that K-Means Clustering could classify small and medium enterprises into three groups based on financial performance, product sales, and marketing strategies. The resulting clusters provided information to support more targeted SME development strategies [12] [13] .

3. Application of K-Means for Sales Data Clustering

K-Means Clustering is an unsupervised machine learning algorithm used to divide a dataset into a predetermined number of clusters based on the distance between data points and cluster centroids. The algorithm works iteratively by determining initial centroids, calculating the distance of each data point to the centroids, assigning data points to the nearest cluster, and updating the centroid values until the clustering process converges. One commonly used distance measurement is Euclidean Distance, which calculates the geometric distance between data points and cluster centers. K-Means is widely used because of its relatively simple implementation, computational efficiency, and ability to identify patterns in numerical datasets. Previous research by Sudrajat et al. (2022) applied K-Means to group SMEs in South Sumatra and demonstrated its ability to classify business data into different clusters. In addition, research by Silitonga et al. (2025) applied K-Means to map small and micro industries across districts and cities in North Sumatra based on business units, workforce, and capital loans, producing three potential categories: low, medium, and high [14].

Research Methodology

This study applies a quantitative research approach using data mining techniques to analyze and map the potential of Small and Medium Industries (SMIs) based on production,

investment, and workforce characteristics. The research uses secondary data obtained from the Small and Medium Industries database for the 2021–2026 period in Tanjungbalai City, North Sumatra, Indonesia. The dataset contains industrial information, including business categories, production types, workforce size, investment value, production capacity, and production value. The data are processed systematically to identify similarities and differences among industrial units through the K-Means Clustering algorithm.

The research stages consist of data collection, data preprocessing, feature selection and normalization, determination of the number of clusters, implementation of the K-Means Clustering algorithm, and cluster evaluation and interpretation. Each stage is designed to ensure that the resulting clusters accurately represent the characteristics of the observed industrial data. The clustering results are expected to classify SMIs into different potential categories, providing useful information for understanding industrial conditions and supporting data-driven decision-making for regional industrial development.

The research process consists of six systematic stages, starting from data collection and ending with the evaluation and interpretation of clustering results. Each stage is carried out sequentially to ensure that the data processing process produces relevant and reliable information regarding the potential of Small and Medium Industries.

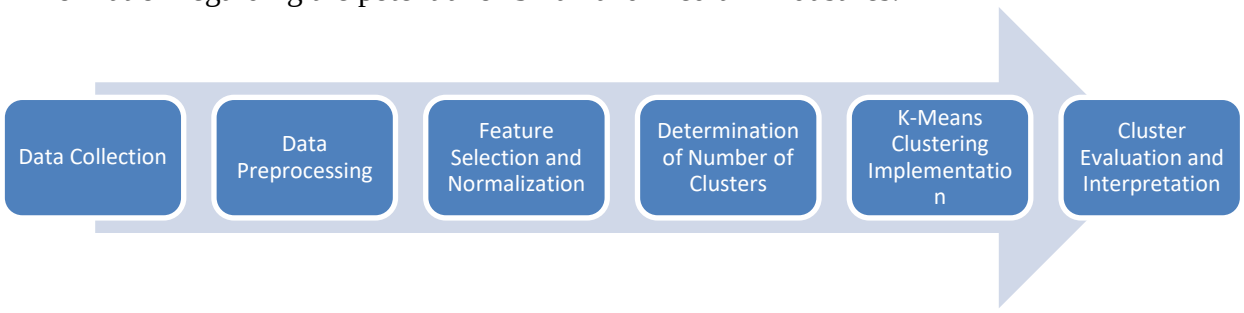


Figure 1. Research Methodology

1. Data Collection
   The first stage involves collecting secondary data on Small and Medium Industries (SMIs) in Tanjungbalai City for the 2021–2026 period. The data contain information about business names, production types, industrial sectors, workforce size, investment value, production capacity, and production value. These data are obtained from the industrial database and serve as the primary source for the research analysis. The collected data are subsequently used to identify the characteristics and potential of each industrial unit.
2. Data Preprocessing
   The second stage involves preparing the collected data to ensure that they are suitable for the clustering process. This process includes checking incomplete records, identifying duplicate data, examining inconsistent data formats, and handling invalid numerical values. Data preprocessing is necessary because the quality and consistency of the dataset can influence the accuracy of the clustering results. Through this stage, the dataset is prepared in a structured format for further analysis.
3. Feature Selection and Normalization
   The third stage focuses on selecting the variables that are relevant to identifying the potential of Small and Medium Industries. The selected variables consist of investment value, workforce size, production capacity, and production value. These variables are considered important because they represent the financial, operational, and production characteristics of each industrial unit. After feature selection, data normalization is performed to ensure that all variables have comparable measurement scales, preventing variables with larger numerical values from dominating the clustering process.
4. Determination of Number of Clusters
   The fourth stage involves determining the appropriate number of clusters for grouping the industrial data. The Elbow Method is used to examine different cluster configurations and

identify the optimal number of groups based on the changes in clustering error. This process helps determine the most suitable cluster structure for representing the characteristics of the industrial dataset. The study considers three potential categories, namely high, medium, and low, with the final classification determined by the clustering results.

5. K-Means Clustering Implementation
   The fifth stage involves implementing the K-Means Clustering algorithm to group industrial units according to similarities in their selected characteristics. The algorithm begins by determining initial cluster centers, assigning each data point to the nearest cluster, and updating the cluster centers based on the assigned data. This process is repeated until the cluster assignments become stable. The resulting groups represent industrial units with similar characteristics in terms of investment, workforce, production capacity, and production value.
6. Cluster Evaluation and Interpretation
   The final stage involves evaluating and interpreting the clustering results to determine the quality and characteristics of each group. The Davies-Bouldin Index (DBI) is used to evaluate clustering performance, where a lower value indicates better cluster separation and compactness. Furthermore, each cluster is analyzed based on its centroid characteristics to identify the differences in investment, workforce, production capacity, and production value. The final results are used to classify industrial potential into high, medium, and low categories and provide information for supporting regional industrial development strategies.

Results

The analysis uses the Small and Medium Industries (SMIs) directory for Tanjungbalai City covering the 2021–2026 data period. The source workbook contains 709 records and provides information on enterprise identity, location, production type, industrial branch, workforce, investment, production capacity, production unit, production value, and data year. For the clustering analysis, 708 records with complete values for workforce, investment, and production value were retained. These three numerical indicators were selected to represent labor input, financial resources, and production output.

The results are organized into three tables and three figures. Table 1 presents the dataset profile and descriptive statistics of the numerical attributes. Table 2 provides the complete distribution of records by data year and normalized industrial branch. Table 3 summarizes the K-Means clusters and their numerical characteristics. The figures visualize the annual data distribution, industrial branch composition, and final cluster proportions.

1. Dataset Overview and Descriptive Statistics

Table 1 summarizes the scope of the source dataset, the completeness of the selected attributes, and the descriptive statistics of the numerical indicators available in the IKM directory.

Table 1. Dataset profile and descriptive statistics of numerical attributes.

Indicator	Result / descriptive statistics
DATASET PROFILE	
Total records in source dataset	709
Data period	2021–2026
Distinct enterprise names recorded	617
Records with complete clustering attributes	708
Records excluded due to missing workforce value	1
Selected clustering attributes	Workforce, investment, production value

Number of clusters in final model	3
NUMERICAL ATTRIBUTE	VALID N | MINIMUM | MEDIAN | MEAN | MAXIMUM
Workforce (persons)	708 | 1.00 | 1.00 | 1.75 | 72.00
Investment (IDR)	709 | 100,000 | 2,415,000 | 18,306,571 | 1,000,000,000
Production capacity (recorded units)	709 | 11.00 | 616.00 | 20,478.36 | 1,375,000.00
Production value (IDR)	709 | 550,000 | 30,000,000 | 197,124,275 | 46,237,400,000

The source dataset contains 709 records and 617 distinct enterprise-name entries. Since some enterprise names may appear in more than one data year, the number of distinct names is not treated as an exact count of active enterprises. The workforce attribute contains one missing value, while investment, production capacity, production value, and data year are available for all records. Therefore, 708 complete records were used in the clustering process. The descriptive statistics also show substantial differences between minimum, median, mean, and maximum values, particularly for investment and production value, indicating a strongly right-skewed distribution.

2. Distribution of IKM Records

Table 2 presents the full distribution of records across the six data years and the industrial branches recorded in the source file. Variants in spelling and capitalization were consolidated for descriptive reporting only.

Table 2. Distribution of IKM records by year and industrial branch

Category	Number of records	Percentage
Distribution By Data Year		
2021	50	7.05%
2022	50	7.05%
2023	129	18.19%
2024	326	45.98%
2025	137	19.32%
2026	17	2.40%
Total by year	709	100.00%
Distribution By Industrial Branch		
Food	335	47.25%
Clothing	269	37.94%
Metal	60	8.46%
Crafts	28	3.95%
Furniture	13	1.83%
Not recorded	4	0.56%
Total by branch	709	100.00%

The annual distribution shows that 2024 has the largest number of records, with 326 records (45.98%), followed by 2025 with 137 records and 2023 with 129 records. The industrial branch distribution is dominated by Food with 335 records, followed by Clothing with 269 records and Metal with 60 records. Minor spelling variants such as capitalization and typographical differences were consolidated into the corresponding branch categories; unrecognized or blank entries remain reported separately as recorded/unclassified. These

distributions describe the composition of the provided directory and should not be interpreted as a complete census of all IKM activity.

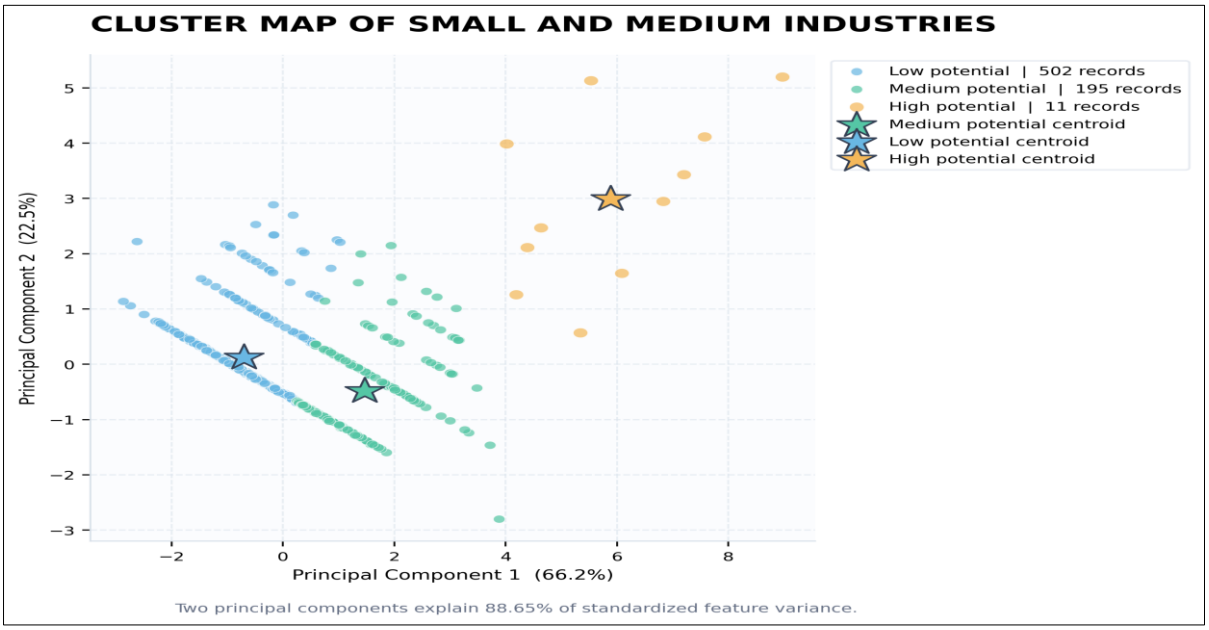


Figure 1. PCA-based scatter plot of IKM records grouped by K-Means clusters.

Figure 1 presents the distribution of IKM records in a two-dimensional space produced by Principal Component Analysis (PCA) from the three standardized clustering attributes. Each point represents one IKM record, while different colors indicate the low-, medium-, and high-potential clusters generated by K-Means. The X-shaped markers represent the cluster centroids projected into the same PCA space. The first two principal components retain 88.65% of the total variance, providing a useful two-dimensional representation of the clustering structure.

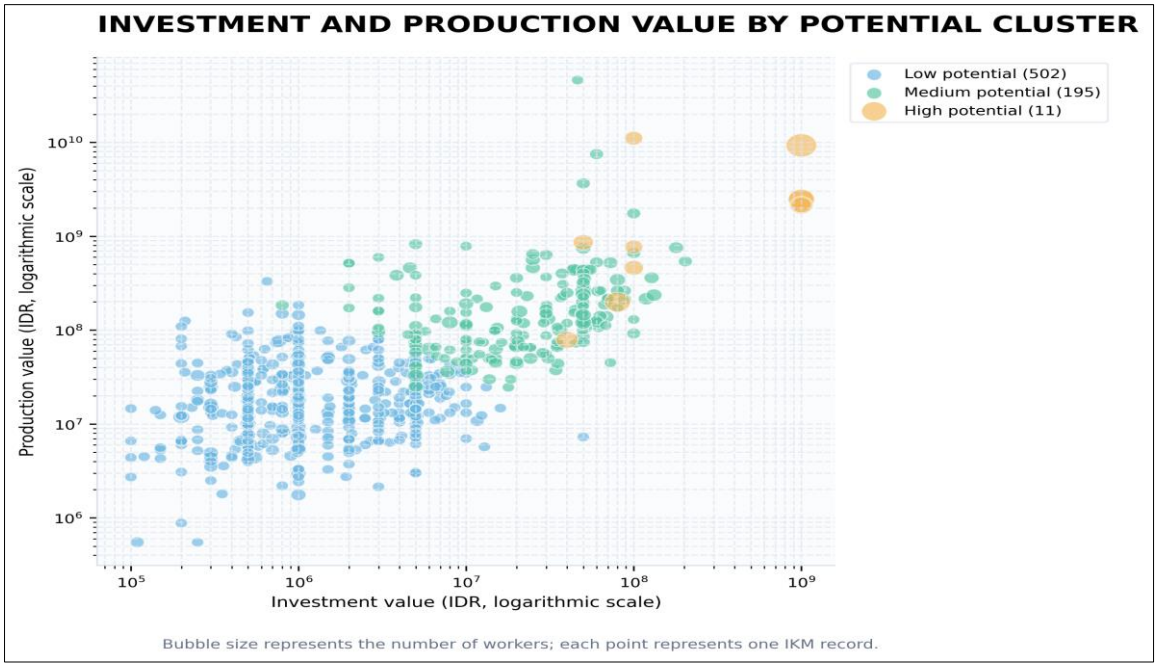


Figure 2. Scatter plot of investment value and production value by K-Means cluster

Figure 2 maps each IKM record according to its investment value and production value. The horizontal and vertical axes use logarithmic scales to make observations with substantially different monetary values visible in the same plot. Colors distinguish the three K-Means potential groups, while X-shaped markers indicate the median investment and median production value within each group. The visualization shows how the clusters differ in their financial input and production output profiles.

3. K-Means Clustering Results

The K-Means model was applied to the three selected numerical attributes: workforce, investment value, and production value. Because these variables have different scales and highly skewed distributions, a logarithmic transformation followed by standardization was applied before model training. Candidate cluster numbers from K = 2 to K = 6 were compared using the Davies-Bouldin Index (DBI).

Table 3. K-Means clustering evaluation and cluster characteristics.

Cluster / model	Records / DBI	Share / Silhouette	Mean workforce	Mean investment / mean production
MODEL EVALUATION				
K = 2	0.930	0.475		
K = 3	0.814	0.436		Selected
K = 4	1.026	0.295		
K = 5	0.956	0.319		
K = 6	0.873	0.334		
CLUSTER PROFILE				
Low potential	502	70.90%	1.23	IDR 2,257,789 / IDR 25,889,426
Medium potential	195	27.54%	1.83	IDR 32,568,970 / IDR 485,319,586
High potential	11	1.55%	23.82	IDR 497,272,727 / IDR 2,917,027,273
Total	708	100.00%	—	—

Among the tested configurations, K = 3 achieved the lowest DBI value (0.814), compared with K = 2, 4, 5, and 6. The three-cluster solution was therefore used for the final grouping. The corresponding Silhouette Coefficient was 0.436. The low-potential cluster contains 502 records (70.90%), with an average workforce of 1.23, average investment of IDR 2,257,789, and average production value of IDR 25,889,426. The medium-potential cluster contains 195 records (27.54%), while the high-potential cluster contains 11 records (1.55%). The high-potential group has the highest average values across the selected attributes, but its small size means that individual records should be reviewed before using the result to prioritize assistance or development programs.

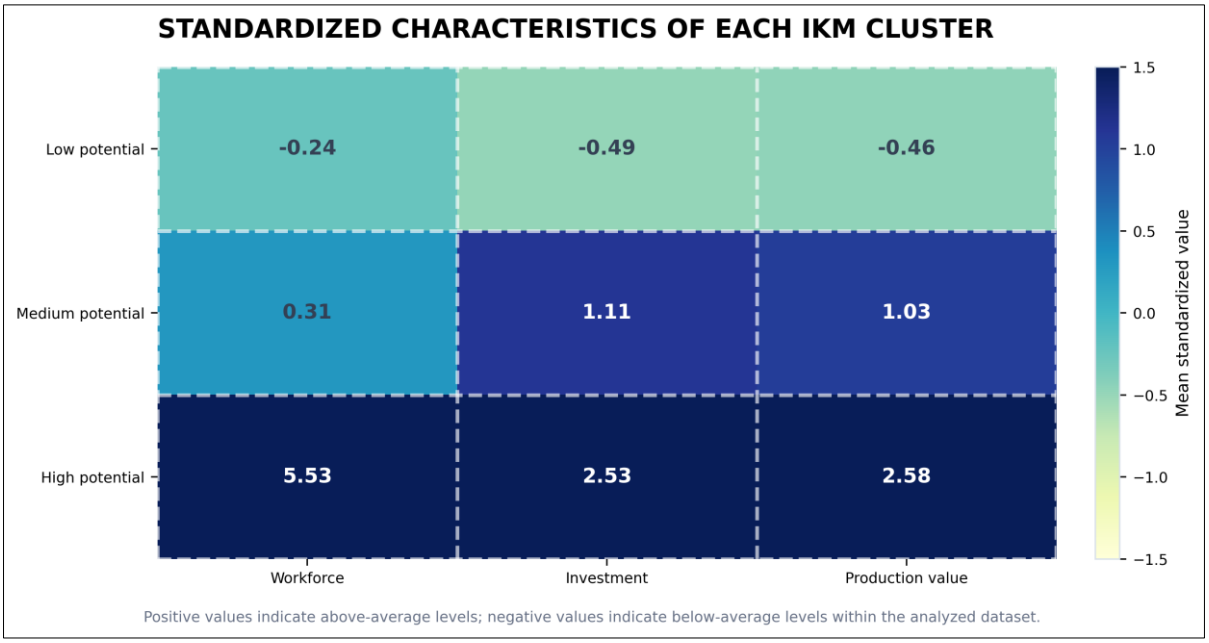


Figure 3. Bubble scatter plot of workforce and production value across K-Means clusters.

Figure 3 illustrates the relationship between workforce size and production value for each IKM record. The color of each point represents its K-Means potential category, while bubble size reflects the relative investment value. The horizontal axis uses a symmetric logarithmic scale to accommodate records with very small or zero workforce values, and the vertical axis uses a logarithmic scale to display the wide range of production values. This visualization makes it possible to compare labor input and production output while retaining investment information through bubble size.

Discussion

The results demonstrate that the K-Means Clustering algorithm can identify and group Small and Medium Industries (SMIs) based on workforce size, investment value, and production value. The analysis of 708 records produced three clusters: low potential consisting of 502 records (70.90%), medium potential consisting of 195 records (27.54%), and high potential consisting of 11 records (1.55%). These findings indicate that most industrial records have relatively low investment and production characteristics, while only a small proportion demonstrate higher workforce involvement and production capacity. The clustering results provide an overview of the differences in industrial characteristics and can help identify groups of businesses with similar development potential.

The model evaluation produced a Davies-Bouldin Index (DBI) of 0.814 and a Silhouette Coefficient of 0.436, indicating that the clustering model identifies distinguishable groups, although some similarities between clusters remain. The PCA visualization explains 88.65% of the total data variance through its first two principal components, supporting the visual interpretation of the clustering structure. These findings can support regional industrial development planning by providing information for prioritizing assistance, improving investment capacity, strengthening workforce development, and increasing production performance. Therefore, K-Means Clustering can serve as an analytical approach to support data-driven decision-making in the development of Small and Medium Industries..

Conclusion

This study demonstrates that the K-Means Clustering algorithm can be applied to map and analyze the potential of Small and Medium Industries (SMIs) based on workforce size, investment value, and production value. Using 708 complete records, the clustering process generated three groups consisting of 502 low-potential records (70.90%), 195 medium-potential records (27.54%), and 11 high-potential records (1.55%). These results reveal differences in industrial characteristics, particularly in terms of investment, workforce involvement, and production performance. The model achieved a Davies-Bouldin Index (DBI) of 0.814, indicating the quality of the resulting cluster structure.

The findings provide useful information for understanding the distribution of SMI potential and supporting data-driven industrial development planning. The clustering results can assist policymakers in identifying industrial groups that require different development priorities, including investment support, workforce improvement, production optimization, and business expansion. Future research is expected to incorporate additional variables, such as market access, business revenue, and production efficiency, to improve the comprehensiveness of industrial potential analysis.

References

[1] H. W. Dhany, F. Izhari, and S. Komputer, "Jurnal ICT : Information and Communication Technologies, 15 (2) (2024) 48-54 Prediction analysis condition animal use algorithm (SVM+KNN)," 2024. [Online]. Available: www.ejournal.marqchainstitute.or.id/index.php/JICT

[2] Z. Sitorus, D. Prayogi, M. A. Rizko, A. G. Suteja, and M. R. Harahap, "Implementation of the Insertion Sort Algorithm to Sort Positive Integers in Ascending Order Using Flowgorithm," *Journal of Information Technology, computer science and Electrical Engineering (JITCSE)*, vol. 1, no. 3, pp. 323–328, 2024, doi: 10.30596/jitcse.

[3] K. Khairul, A. H. Nasyuha, A. Ikhwan, M. H. Aly, and A. Ahyanuardi, "Implementation of Multiple Linear Regression to Estimate Profit on Sales of Screen Printing Equipment," *JURNAL INFOTEL*, vol. 15, no. 2, pp. 55–61, Jun. 2023, doi: 10.20895/infotel.v15i2.934.

[4] M. S. Novelan and S. Aryza, "OPTIMIZATION CVRP WITH MACHINE LEARNING FOR IMPROVED CLASSIFICATION OF IMBALANCED DATA FOOD DISTRIBUTION," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 4, pp. 917–925, Jun. 2025, doi: 10.33480/jitk.v10i4.6467.

[5] R. Farta Wijaya, F. Kurniawan, R. R. Putrai, A. Alvin, F. Sains, and D. Teknologi, "OPTIMIZING THE MANAGEMENT OF VILLAGE ACTIVITIES THROUGH INFORMATION SYSTEMS FOR TRANSPARENCY AND ACCOUNTABILITY IN PERTUMBUKAN VILLAGE, WAMPU DISTRICT," 2024.

[6] M. M. Maulana, A. I. Suroso, Y. Nurhadryani, and K. B. Seminar, "The Smart Governance Framework and Enterprise System's Capability for Improving Bio-Business Licensing Services," *Informatics*, vol. 10, no. 2, Jun. 2023, doi: 10.3390/informatics10020053.

[7] A. Putera Utama Siahaan, N. Azizah Harahap, R. Yuni Simanullang, and P. Wanny, "Analysis of Inpatient Data Using Cluster Analysis on Simulation Dataset," *Bulletin of Information Technology (BIT)*, vol. 6, no. 1, pp. 33–39, 2025, doi: 10.47065/bit.v5i2.1830.

[8] M. S. Novelan, S. Efendi, P. Sihombing, and H. Mawengkang, "VEHICLE ROUTING PROBLEM OPTIMIZATION WITH MACHINE LEARNING IN IMBALANCED CLASSIFICATION VEHICLE ROUTE DATA," *Eastern-European Journal of*

Enterprise Technologies, vol. 5, no. 3(125), pp. 49–56, 2023, doi: 10.15587/1729-4061.2023.288280.

[9] M. Iqbal and S. Efendi, "Data-Driven Approach for Credit Risk Analysis Using C4.5 Algorithm," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 14, no. 1, pp. 11–20, May 2023, doi: 10.21512/comtech.v14i1.8243.

[10] J. Asharf, N. Moustafa, H. Khurshid, E. Debie, W. Haider, and A. Wahab, "A review of intrusion detection systems using machine and deep learning in internet of things: Challenges, solutions and future directions," Jul. 01, 2020, *MDPI AG*. doi: 10.3390/electronics9071177.

[11] W. Cetera, W. Gogołek, A. Żołnierski, and D. Jaruga, "Potential for the use of large unstructured data resources by public innovation support institutions," *J. Big Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00610-6.

[12] G. Rajeshkumar *et al.*, "Smart office automation via faster R-CNN based face recognition and internet of things," *Measurement: Sensors*, vol. 27, Jun. 2023, doi: 10.1016/j.measen.2023.100719.

[13] A. Rejeb *et al.*, "The Internet of Things (IoT) in healthcare: Taking stock and moving forward," Jul. 01, 2023, *Elsevier B.V.* doi: 10.1016/j.iot.2023.100721.

[14] M. Overton, S. Larson, L. J. Carlson, and S. Kleinschmit, "Public data primacy: the changing landscape of public service delivery as big data gets bigger," *Global Public Policy and Governance*, vol. 2, no. 4, pp. 381–399, Dec. 2022, doi: 10.1007/s43508-022-00052-z.

[15] J. Ilmiah, E. & Erudisi, A. Zahara, M. I. Kabullah, and R. E. Putera, "JIEE: Effectiveness of the OSSRBA (Online Single Submission Risk Based Approach) System in Business Licensing Services in Payakumbuh City DPMPTSP," vol. 3, pp. 22–40, 2023.

[16] W. Jin, "Challenges and Innovative Countermeasures Faced by Public Administration in the Context of Big Data and Internet of Things," *Math. Probl. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/8949365.