A Hybrid Approach Using K-Means Clustering and the SAW Method for Evaluating and Determining the Priority of SMEs in Palembang City

Terttiaavini[1]
[1]Master of Computer Science, Faculty of Computer Science and Sciences, Indo Global Mandiri University, Indonesia

Corresponding author: Terttiaavini (avini.saputra@uigm.ac.id)

ABSTRACT The current efforts to develop Small and Medium Enterprises (SMEs) are still facing challenges in setting appropriate targets. Although the Palembang City Cooperative and SME Agency has launched various programs and initiatives to support SME development, they have not yet successfully identified the SMEs that should be given priority for development. This study aims to apply a hybrid approach that combines the K-Means Clustering method and Simple Additive Weighting (SAW) to evaluate and prioritize SME development in Palembang City. The K-Means Clustering method is used to group SMEs based on their characteristics, while SAW provides preference values (V_i). The SME data was obtained from the Palembang City Cooperative and SME Agency, covering 362 SME units. The K-Means Clustering results yielded two clusters: Cluster 0 as the High Growth Cluster and Cluster 1 as the Stability and Improvement Cluster. Validation using cross-validation showed that this model achieved an accuracy of 99.72%. The SAW analysis on Cluster 0 indicated that the Kopi Kaljo SME received the highest priority with a preference value of 45.71. This study confirms that this hybrid approach is effective in grouping SMEs based on their characteristics and prioritizing them based on data-driven evaluation. The research results are expected to help the Palembang City Cooperative and SME Agency design more effective and targeted assistance programs to optimize the contribution of SMEs to local economic growth to the maximum extent.

KEYWORDS Hybrid approach, K-Means Clustering, Simple Additive Weighting, SMEs

I. INTRODUCTION

Small and Medium Enterprises (SMEs) play a crucial role in the local and national economy, especially in developing cities like Palembang. SMEs not only create employment opportunities but also contribute to overall economic growth. In Palembang, SMEs serve as the main driver in improving community welfare and reducing unemployment rates. They also act as a pillar in strengthening the local economic structure, providing stability in times of economic crisis [1].

The Cooperative and SMEs Office of Palembang City has launched various programs and initiatives to support the development of SMEs in the region. However, the main challenge is how to classify SMEs based on certain characteristics and how to effectively prioritize their development. The data from the Department of Cooperatives and SMEs of Palembang City in 2022 recorded 1,103 SMEs, while more than 160,000 SMEs remain unregistered [2]

Traditional methods often fall short in handling the complexity of diverse SMEs data. With various types of SMEs having different characteristics and performances, an appropriate approach is needed to evaluate and determine their development priorities effectively [3].

One of the main issues faced by policymakers and SME managers is how to classify SMEs based on certain characteristics and how to effectively prioritize their development [4]. A more accurate and data-driven approach is needed to evaluate the performance and potential of SMEs more precisely, thereby providing targeted and appropriate support according to the needs of each SME [5].

This study aims to develop and implement a hybrid approach combining K-Means Clustering and Simple Additive Weighting (SAW) methods to evaluate and determine the development priorities of SMEs in Palembang City [6]. The K-Means Clustering method is used to group SMEs based on their characteristics, while the SAW method is employed to assign preference values to each clustered

SME, thus identifying those with the highest performance and potential [7].

The method used in this study is a hybrid approach combining two data mining techniques. First, K-Means Clustering is used to group SMEs based on their characteristics [8]. This approach helps identify underlying patterns among different SME groups. Second, Simple Additive Weighting (SAW) is used to assign preference values to each clustered SME, allowing the identification of SMEs with the highest performance and potential within each cluster. The combination of these two techniques is expected to provide a more comprehensive and accurate overview of the SME conditions in Palembang City and support better decision-making in the development and support of SMEs [9].

The benefits of this study include a deeper understanding of SMEs in Palembang City by identifying the underlying patterns and characteristics of various SME groups. By using the K-Means Clustering and Simple Additive Weighting (SAW) approaches, this study provides development priority recommendations for SMEs based on more accurate data-driven evaluations. This is expected to enhance the effectiveness of support for SMEs through the design of more adaptive and sustainable assistance programs, thereby maximizing the contribution of SMEs to local economic growth [10][11].

By implementing this hybrid model, it is expected to provide strategic recommendations to the Department of Cooperatives and SMEs of Palembang City in designing more effective and targeted assistance programs for priority SMEs. Through comprehensive and data-driven evaluation, this model enables more accurate identification of the needs and potentials of SMEs within each cluster. This is anticipated to enhance the efficiency of resource allocation and assistance, and strengthen the contribution of SMEs to sustainable local economic growth [12].

II. LITERATURE REVIEW

A literature review of the K-Means Clustering method, Simple Additive Weighting (SAW), and the application of RapidMiner can provide an in-depth understanding of the concepts, applications, and relevance of each method in the context of the development and evaluation of Small, and Medium Enterprises (SMEs). Here is an overview of the literature review used:

A. K-MEANS CLUSTERING METHOD

The K-Means Clustering method is a data analysis technique used to group data into different clusters based on certain similarities in characteristics [13]. This technique is widely applied in various studies due to its effectiveness in clustering data without prior labels or supervision [14]. The stages in the K-Means Clustering method are as follows:

1. Centroid Initialization
   Randomly select K initial centroids from the data points as the initial cluster centers.
2. Data Point Allocation to Clusters
   Assign each data point to the nearest cluster based on the Euclidean distance between the data point and the centroid.
3. New Centroid Calculation
   Recalculate the position of the new centroid in each cluster by taking the average of all data points that belong to the cluster.
4. Iteration
   Repeat steps 2 and 3 until a stopping condition is met, such as no significant changes in the centroid positions or the maximum number of iterations is reached.

Formulas Used in K-Means Clustering

1. Euclidean Distance
   To calculate the distance between two points in n-dimensional space, Equation (1) is the Euclidean formula [15].

$$distance(X_i, C_j) = \sqrt{\sum_{k=1}^{n} \left(X_{ik} - C_{jk}\right)^2} \tag{1}$$

   X_i is the i-th data point, C_j is the j-th centroid, and n is the number of dimensions.
2. Centroid Update
   After all data points are allocated to clusters, the new centroid C_j is calculated as the average of all data points X_i that belong to the j-th cluster. Equation (2) represents the formula for calculating the new centroid C_j.

$$C_j = \frac{1}{|S_j|} \sum_{X_i \in S_j} X_i \tag{2}$$

   S_j is the number of data points in the j-th cluster.

The use of the K-Means algorithm for clustering SMEs offers advantages in identifying patterns in data without the need for prior class labels. This algorithm is scalable for large datasets, easy to interpret, and aids in determining the optimal clusters using the Elbow method [16]. By utilizing the Elbow method, the K-Means algorithm can automatically determine the optimal number of clusters based on significant drops in the Sum of Squared Errors (SSE) values. This enables researchers or practitioners to efficiently and accurately group SMEs based on data characteristics.

B. THE SIMPLE ADDITIVE WEIGHTING (SAW) METHOD

SAW is a multi-criteria decision-making technique used to evaluate alternatives based on the relative weights of each criterion [17]. SAW is employed to assign preferences to SMEs that have been clustered using the K-Means Clustering method. With SAW, each SME is assessed according to several predetermined criteria. The steps involved in the SAW method are as follows:

1. Determining Criteria (C_i)
   Criteria selected should be relevant and representative of the evaluation goals. These criteria are typically chosen based on an analysis of the needs and characteristics of the SMEs being evaluated.
2. Determining Suitability Ratings (R) and Weights (W_i)
   Each alternative A_i is assessed using suitability ratings R_{ij} for each criterion C_j. Weights W_j are assigned to indicate the relative importance of each criterion $\boldsymbol{C_j}$.
3. Creating the Decision Matrix (X) and Normalization
   The decision matrix X has dimensions $m \times n$, where m is the number of alternatives and n is the number of criteria. Each element X_{ij} of matrix $\boldsymbol{X}$ represents the suitability rating R_{ij} of alternative A_i for criterion C_j.

$$X = \begin{bmatrix} X_{11} & \dots & X_{1n} \\ \dots & \dots & \dots \\ X_{m1} & \dots & X_{mn} \end{bmatrix}$$

   This structured approach allows for a systematic evaluation of SMEs based on weighted criteria, facilitating informed decision-making in developmental and support programs.
4. Normalization of the Decision Matrix (R)
   Normalization is performed to transform each element X_{ij} into the same range, based on whether the attribute is a benefit or a cost attribute. The normalization of the decision matrix is done using (3) and (4):

   For benefit attributes

$$r_{ij} = \frac{X_{ij}}{\max X_{ij}} \quad (3)$$

   For cost attributes

$$r_{ij} = \frac{\min X_{ij}}{X_{ij}} \quad (4)$$

   r_{ij} is the normalized value of element i on attribute j, where X_{ij} is the original value of element i on attribute j. $\max X_{ij}$ represents the maximum value of attribute j, and $\min X_{ij}$ represents the minimum value of attribute j. Normalization can be performed using various methods, such as min-max normalization or z-score normalization, depending on the nature of the data.
5. The calculation of Preference Value (V_i)
   After obtaining the normalized matrix R, the preference value V_i for each alternative A_i is calculated by summing the products of matrix R with the weight vector W using (5).

$$V_i = \sum_{j=1}^{n} W_j \times R_{ij} \quad (5)$$

   V_i is the preference value or score for alternative A_i

   The final result of the SAW process is the ranking of alternatives based on the value of V_i. Alternatives with higher V_i values are considered the best solutions or highest priorities

C. RAPIDMINER APPLICATION

RapidMiner is an open-source platform that provides various tools for data analysis, including data mining processes, predictive modeling, and business analytics. RapidMiner can help optimize the evaluation and decision-making processes related to SMEs by leveraging its visualization tools, data processing capabilities, and modeling functionalities offered by the platform [18].

By integrating literature on this topic, the research can develop a holistic approach to evaluating and developing SMEs using K-Means Clustering and SAW with the assistance of RapidMiner. This literature review will provide a strong theoretical foundation and practical insights to design effective and applicable research methodologies in the context of SMEs in Palembang or other regions.

D. HYBRID APPROACH FOR EVALUATING AND PRIORITIZING SMEs

This research aims to address existing research gaps by introducing a novel hybrid approach that combines the K-Means Clustering and Simple Additive Weighting (SAW) methods. The existing research gap lies in the challenge of prioritizing and developing Small and Medium Enterprises (SMEs) based on their diverse characteristics and needs.

Previous studies have contributed by categorizing SMEs into various clusters such as high, medium, and low [19], independent, developing, and assisted [20], as well as micro and small businesses [21], and strong and weak sustainability groups [22]. However, their weakness lies in their limited focus solely on classification and categorization. The approaches used tend to be descriptive and lack the utilization of objective data to determine development priorities. This limitation restricts the ability to provide specific and strategic recommendations for SMEs.

Traditional methods often struggle to manage the complexity and variation present in SME data, making it difficult to determine which SMEs should receive priority support and development.

Here is a detailed explanation of how this approach innovates and adds value compared to existing methods:

1. Hybrid Approach: The integration of K-Means Clustering allows segmentation of SMEs into different groups based on their characteristics such as income, number of employees, and business scale. This clustering provides a fundamental understanding of SME clusters, identifying groups like "High Growth" and "Stability and Improvement," which represent SMEs with different development needs and potentials.
2. SAW Method: After clustering, the SAW method is used to assign preference values (V_i) to each SME within

the identified clusters. This method evaluates SMEs based on predefined criteria to objectively measure development priorities.
3. Comprehensive Evaluation: Unlike traditional subjective approaches, this hybrid model ensures comprehensive, data-driven evaluation of SMEs. It leverages statistical analysis and machine learning techniques to gain insights from a dataset encompassing 362 SMEs from the Cooperative and SME Agency of Palembang City.
4. Value Proposition: The innovation lies in seamlessly integrating clustering for segmentation and SAW for prioritization, enabling policymakers and stakeholders to design assistance programs tailored to the identified needs of SME clusters. This approach optimizes resource allocation and enhances the effectiveness of support programs, thereby maximizing SME contributions to sustainable local economic growth.
5. Comparison with Existing Methods: Unlike single-method approaches that may overlook nuanced differences among SMEs or rely solely on subjective evaluations, the hybrid model in this study offers a structured and objective framework. It combines the strengths of clustering (for grouping similar SMEs) and SAW (for prioritizing based on criteria) to provide a holistic view that traditional methods may lack.

Overall, this research aims to bridge gaps by introducing a hybrid approach that is effective not only in categorizing SMEs but also in prioritizing them based on objective criteria. This innovation is expected to improve the accuracy and effectiveness of policy formulation and strategic planning for SME development in Palembang, offering a model that can be applied and adapted in similar contexts.

III. RESEARCH METHODOLOGY

In this section, a detailed explanation will be provided regarding the steps and approaches used to implement the K-Means Clustering and Simple Additive Weighting (SAW) methods in evaluating and prioritizing the development of Small, and Medium Enterprises (SMEs) in Palembang City [23]. Figure 1 illustrates the research stages, covering the process from start to finish in implementing the hybrid approach using K-Means Clustering and Simple Additive Weighting (SAW).

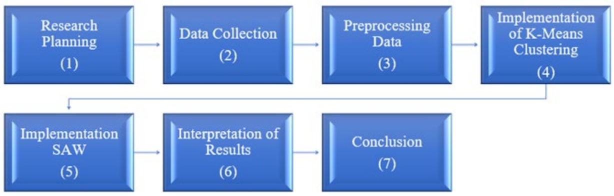

FIGURE 1. Research Stages

To achieve the objectives of this research, several stages will be detailed comprehensively. These stages are designed to ensure that the research is conducted systematically and comprehensively, so that the results obtained can significantly contribute to the evaluation and development of SMEs in Palembang. The following are the research stages to be implemented

A. DATA COLLECTION

Data collection for this research utilizes information provided by the Department of Cooperatives and SMEs of Palembang City. This data includes details from approximately 362 Small, and Medium Enterprises (SMEs) operating in Palembang. Sourcing data from this department is considered highly relevant as it provides direct access to information on characteristics, financial performance, and other factors influencing SMEs in the region [24]. The acquired data includes the SME name, owner's name, education, ownership status, business location status, district, business scale, business type, number of employees, revenue, operational costs, profit, average production quantity, buyer category, target customers, products, monthly sales volume, sales method, and transaction method. The total dataset consists of 362 items.

B. DATA PREPROCESSING

Data preprocessing is a crucial stage in the data analysis process aimed at cleaning, organizing, and preparing raw data for further analysis. In this stage, RapidMiner application is used for data preprocessing. The explanation of the data preprocessing stage is as follows:
1. Data Cleaning This process involves examining the data, determining attributes, and handling missing or incomplete values. Based on statistical analysis using RapidMiner, all data is complete, without outliers, and ready for further analysis.
2. Feature Selection Next, relevant and significant features are selected for clustering analysis and evaluation using SAW. The selected features include Education, District, Business Scale, Number of Employees, Revenue, Operational Costs, Profit, Average Sales, Number of Products Sold, and Transaction Method. Table I shows the features in Palembang City SMEs
3. Data Transformation
In this stage, data is converted and adjusted to prepare it for further analysis. The data transformation process involves converting categorical data into numerical values, specifically for Education, Business Scale, and Transaction Method.
4. Data Normalization
In the context of SME data, the range of values for each criterion can vary significantly. Data normalization aims to standardize the scale of input variables so that different ranges of values do not distort the results of clustering. Through normalization, variables with larger scales do not dominate the distance calculation between

data points, thus preventing bias in cluster formation based on Euclidean distance or other metrics.

TABLE I
FEATURES OF SMEs IN PALEMBANG CITY (*IN THOUSANDS)

Features	data
Business owner's education	High School = 174; Diploma = 42; Bachelor's Degree= 146
District	Alang-alang Lebar = 27; Bukit Kecil = 3; Gandus = 6; Ilir Barat I = 32; Ilir Barat II = 24; Ilir Timur I = 24; Ilir Timur II = 15; Ilir Timur III = 16; Jakabaring = 18; Kalidoni = 24; Kemuning = 14; Kertapati = 3; Plaju = 5; Sako = 38; Seberang Ulu I = 10; Seberang Ulu II = 18; Sematang Borang = 8; Sukajadi Timur = 1; Sukarami = 37; Luar Palembang = 40;
Business Scale	Micro Enterprises = 343; Small Business = 16; Medium Business = 1
Number of Employees*	m = 0 : 55 ; $1 \leq m \leq 3$ = 279 ; $4 < m \leq 6$ = 24; $7 < m \leq 10$ = 3; m = 50 : 1
Revenue	$n \leq 1.000$ = 74; $1.000 < n \leq 5.000$ =168; $5.000 < n \leq 10.000$ = 69; $10.000 < n \leq 50.000$ = 50; n > 100.000 = 1
Operational Costs*	$r \leq 1.000$ =184; $1.000 < r \leq 5.000$ = 141; $5.000 < r \leq 10.000$ =27; r > 10.000 =10
Profit*	$p \leq 1.000$ = 23; $1.000 < p \leq 5.000$ =194; $5.000 < p \leq 10.000$ = 34; $10.000 < p \leq 50.000$ = 11
Average Sales*	$x \leq 10$ =117; $10 < x \leq 50$ = 88; $50 < x \leq 100$ = 42; $100 < x \leq 500$ = 46; $500 < x \leq 1000$ = 28; x > 1000 = 41
Number of Products Sold*	$y \leq 10$ = 120; $10 < y \leq 50$ = 22; $50 < y \leq 100$ = 59; $100 < y \leq 500$ = 87; $500 < y \leq 1000$ = 39; y > 1000 = 35
Transaction Method*	Online = 53; Offline = 51; Both = 258

C. K-MEANS CLUSTERING IMPLEMENTATION

1. Implementing K-Means Clustering using RapidMiner involves several operators: read excel, select attributes, normalize, k-means clustering, and cluster distance performance. Each operator plays a role in preparing and analysing the data to cluster SMEs based on their characteristics. Figure 2 shows the clustering workflow for building a K-Means model using RapidMiner.

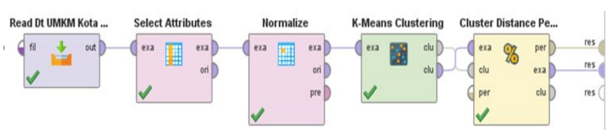

FIGURE 2. Clustering workflow

2. The first step involves setting the number of clusters and The first stage involves setting the number of clusters and relevant attributes using the Select Attributes, Set Parameters, and Normalize operators. Next, the K-Means algorithm will iterate to find the centroid for each cluster and group the data based on its proximity to the centroid using the K-Means operator. The clustering results are evaluated to measure their quality, often by considering the inertia value of the clusters using the Cluster Distance Performance operator, as well as visualizing the patterns formed using the Scatter Plot operator, thus providing valuable insights for decision-making related to the development strategy of SMEs in Palembang City
3. Elbow Method. The Elbow Method is used to determine the optimal number of clusters in cluster analysis. This method involves plotting the Sum of Squared Errors (SSE) values for various numbers of clusters (K) and then identifying the point where the decrease in SSE starts to slow significantly. This point resembles an elbow shape on the plot and indicates the optimal number of clusters for analysis. Table II presents a comparison of centroid distance values for each cluster.

TABLE II
CALCULATE THE ELBOW PLOT IN CLUSTERING

K	Elbow
2	4.925
3	3.927
4	3.234
5	2.262
6	1.694
7	1.239

To visualize the centroid distance values to determine the optimal number of clusters using the elbow method, a line diagram can be displayed as shown in Figure 3.

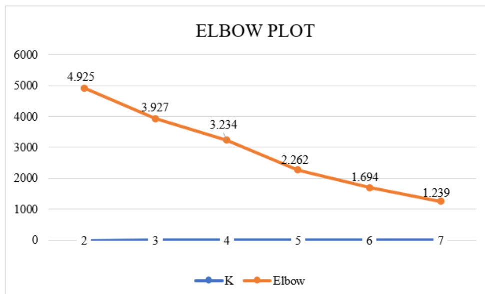

FIGURE 3. Elbow Plot Visualization

The results of the Elbow Plot calculation indicate that five clusters (K=5) are the optimal choice. However, for this analysis, it was decided to use two clusters (K=2).

This decision is based on the research objectives and the consideration of a clearer and more coherent interpretation of the results. The cluster names that represent these two clusters reflect their main characteristics and purposes. The clustering results with K=2 show that Cluster 0, which is the High Growth Cluster, has 314 items, while Cluster 1, which is the Stability and Improvement Cluster, has 48 items, making a total of 362 items. The clustering results with K=2 can be visualized in a Scatter Plot diagram. A Scatter Plot diagram is used to display the relationship between two variables in bivariate data. This patterns, correlations, and trends between these variables diagram helps in analyzing and visualizing data to find. Figure 4 shows the Scatter Plot trends between these variables diagram helps in analyzing and visualizing data to find. Figure 4 shows the Scatter Plot diagram for K=2, which displays the distribution of data within the two formed clusters.

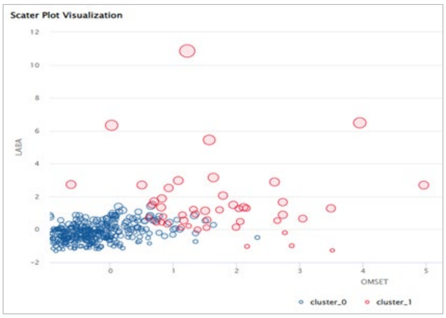

FIGURE 4. Scatter Plot Diagram for K=2

A clearer understanding is needed to measure how well the model can generalize to new data not seen during training. This technique divides the data into several subsets, training the model on one subset and testing it on another subset in turn. This approach provides a more accurate evaluation of the model's performance when faced with unseen data. Cross-validation calculations can be used to determine how well the model can generalize to new data not seen during training, thus providing a more accurate assessment of model performance.

The results of the cross-validation calculation generate a Performance Vector Table. The Performance Vector Table displays performance metrics for each fold used in the cross-validation, along with the mean of these metrics. Performance metrics can include Accuracy, Precision, and Recall values. Table III shows the Performance Vector Table resulting from the cross-validation calculation.

TABLE III
PERFORMANCE VECTOR TABLE

	true cluster_0	true cluster_1	class precision
pred. cluster_0	335	1	99.70%
pred. cluster_1	0	26	100.00%
class recall	100.00%	96.30%	

The cross-validation test results are detailed as follows:

1. Overall Accuracy: 99.72% This accuracy value indicates that the model can cluster SMEs with a success rate of 99.72% of the total data tested. This very high accuracy level demonstrates the model's excellent ability to distinguish between clusters.
2. Pred. Cluster_0**:** A total of 335 data points that belong to Cluster_0 was correctly grouped into Pred. Cluster_0. Only 1 data point from True Cluster_1 was incorrectly grouped into Pred. Cluster_0. The precision for Pred. Cluster_0 is 99.70%, meaning that of all the data predicted as Cluster_0, 99.70% actually belong to Cluster_0.
3. Pred. Cluster_1: All data points that belong to Cluster_1 (26 data points) was correctly grouped into Pred. Cluster_1. No data from True Cluster_0 was incorrectly grouped into Pred. Cluster_1. The precision for Pred. Cluster_1 is 100.00%, meaning that all data predicted as Cluster_1 actually belong to Cluster_1.
4. Class Recall**:** The recall for Cluster_0 is 100.00%, meaning all data points that should belong to Cluster_0 was correctly grouped. The recall for Cluster_1 is 96.30%, meaning that of all data points that should belong to Cluster_1, 96.30% were correctly grouped, with an error rate of only 3.70%.

Overall, these test results show that the implemented clustering model is highly reliable and can be used with a high degree of confidence to cluster SMEs in Palembang according to the specified characteristics. This provides confidence that the clustering results can be used as a reference in designing more targeted assistance and development programs.

D. IMPLEMENTATION OF SIMPLE ADDITIVE WEIGHTING (SAW)

After conducting cross-validation testing and identifying that Cluster 0 has relevant results for further processing, the data will be ranked to help determine the most optimal SMEs. Ranking is performed using the Simple Additive Weighting (SAW) method. The criteria are determined based on the previously established criteria. The steps in the SAW method are explained as follows:

1. Determination of Criteria (C_i)
   Based on the data in Cluster 0 and the SME data, relevant criteria are established to determine the development priorities of the SMEs. These criteria include various aspects such as business owner's education (C1), number of employees (C2), revenue (C3), operational costs (C4), profit (C5), average sales (C6), number of products sold per month (C7), and lending method (C8).
2. Determining Suitability Ratings (R) and Weights (W_i)
   Each alternative (SME) is evaluated or given suitability ratings based on the established criteria. Next, the relative weight of each criterion is determined to establish the importance of each criterion in the decision-making process. The criterion weights are determined based on mathematical analysis. These factors help establish the importance of each criterion in the context of decision-making. The determination of criterion weights (W_i) is explained in Table IV.
3. Normalization of Decision Matrix
   Normalization of the Decision Matrix is performed to ensure that attribute values within the decision matrix are on a uniform scale. This is crucial to ensure fair comparison of each attribute, avoiding bias due to differences in scale and units across different attributes. Through normalization, each attribute is evaluated within the same range, typically [0, 1], thereby making the total score calculation in the SAW method more

accurate and representative. Normalization of the Decision Matrix is conducted using (3) and (4).

TABLE IV
DETERMINATION OF ALTERNATIVES, CRITERIA AND SUITABILITY CHAIN (*IN THOUSANDS)

Criteria (C_i)	Alternatives (A_i)	Suitability Ratings	Weights (W_i)
C1	Business owner's education	High School Diploma Bachelor's Degree	1 2 3
C2	Number of Employees	$0 \leq n < 10$ $10 \leq n < 50$ $n \geq 50$	1 2 3
C3	Revenue*	$0 \leq m < 1$ $1.000 \leq m < 5.000$ $5.000 \leq m < 10.000$ $m \geq 10.000$	1 2 3 4
C4	Operational Costs*	$0 \leq c < 1.000$ $1.000 \leq n < 5.000$ $5.000 \leq n < 10.000$ $n \geq 10.000$	1 2 3 4
C5	Profit*	$0 \leq l < 1.000$ $1.000 \leq l < 5.000$ $5.000 \leq l < 10.000$ $l \geq 10.000$	1 2 3 4
C6	Average Sales	$0 \leq x < 100$ $100 \leq x < 500$ $500 \leq x < 1.000$ $x \geq 1.000$	1 2 3 4
C7	Number of Products Sold	$0 \leq x < 100$ $100 \leq x < 500$ $500 \leq x < 1.000$ $x \geq 1.000$	1 2 3 4
C8	Transaction Method	online offline both	1 2 3

4. Calculation of Preference Value (V_i)
   Calculation of the preference value (V_i) is used to determine the ranking of each alternative based on the predefined criteria. The preference value is computed using equation (5). In this study, the results of the preference value calculation are displayed for the top 10 rankings only. Table V presents the Calculation of Preference Value (V_i) with the top 10 entries.

TABLE V
RESULTS OF PREFERENCE VALUE CALCULATION (V_i)

Initial SMEs	SMEs Name	V_i	Ranking
X208	Kopi kaljo	45,71	1
X187	Warung Neknang	41,31	2
X126	Pempek Ce' Anie	40,79	3
X125	Habar Jumputan	39,45	4
X181	Ikan bakar gegana	37,23	5
X242	Tiara bakery	35,10	6
X313	Benawa Coffee Roastery	35,05	7
X90	Dewul	33,78	8
X150	Rusnani	33,30	9
X225	Kemcum	32,92	10

The result of using the Simple Additive Weighting (SAW) method to determine the most optimal UKM shows that Kopi Kaljo has the highest preference value with a value of 45.71. Kopi Kaljo ranks first in the ranking list, followed by Warung Neknang with a preference value of 41.31. Meanwhile, the lowest value is Arasshop with a value of 10.41.

IV. CONCLUSION

Based on the analysis using the hybrid approach of K-Means Clustering and Simple Additive Weighting (SAW) on SME data in Palembang City, several key conclusions can be drawn:

1. The use of the K-Means Clustering model successfully grouped SMEs into two main clusters: the High Growth Cluster dominated by 314 SMEs, and the Stability with Improvement Cluster consisting of 48 SMEs. This result provides a clear picture of SME distribution based on their characteristics and performance in this region.
2. Validation results of the model showed a very high accuracy rate of 99.72%. The Performance Vector Table confirms that the model effectively classifies SMEs into the appropriate clusters. The High Growth Cluster has a precision of 99.70% and recall of 100.00%, while the Stability with Improvement Cluster has a precision of 100.00% and recall of 96.30%. This indicates that this clustering model is reliable for decision-making related to SME development strategies.
3. The application of the SAW method on clustered SMEs can identify the most optimal SMEs to prioritize in development programs. For instance, SMEs like Kopi Kaljo received the highest preference value with Vi of 45.71, placing it as the top priority for development. This approach allows for a more focused and comprehensive assessment of each SME, ensuring more effective and strategic resource allocation.
4. This research not only provides deep insights into the conditions of SMEs in Palembang City, but also establishes a strong foundation for better decision-making to support local economic growth through more measured and sustainable assistance programs. The implementation of this hybrid model is expected to serve as a valuable guide for stakeholders in designing more effective and supportive policies for SMEs amidst complex economic dynamics.
5. This hybrid method can be further developed by considering the integration of other clustering methods or applying more advanced weighting methods for priority evaluation. Future research could explore how the use of other machine learning techniques like random forest or neural networks could enhance the accuracy and relevance of evaluation results.
6. The implications of these findings are that stakeholders can optimize the type and amount of support provided to SMEs, including training, working capital, and other supportive infrastructure, by understanding their clusters.

AUTHORS CONTRIBUTION

Terttiaavini: Research Design and Conceptualization, Data Collection. Data Preprocessing, Implementation of K-Means Clustering, Implementation of SAW Method, Analysis and Interpretation, Writing and Documentation,Visualization, Review and Editing, Validation and Cross-Validation

COPYRIGHT

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

REFERENCES

[1] T. Tambunan, “Micro, small and medium enterprises in times of crisis: Evidence from Indonesia,” *J. Int. Counc. Small Bus.*, vol. 2, no. 4, pp. 278–302, 2021, doi: 10.1080/26437015.2021.1934754.

[2] Y. Abdullah and I. Gultom, “Dinas Koperasi dan UMKM Palembang bentuk tim data usaha kecil,” *Antara News*. https://sumsel.antaranews.com/berita/643913/dinas-koperasi-dan-umkm-palembang-bentuk-tim-data-usaha-kecil (accessed Apr. 12, 2023).

[3] A. K. M. H. Islam, M. R. Sarker, M. I. Hossain, K. Ali, and K. M. A. Noor, “Challenges of Small- and Medium-Sized Enterprises (SMEs) in Business Growth: A Case of Footwear Industry,” *J. Oper. Strateg. Plan.*, vol. 4, no. 1, pp. 119–143, 2021, doi: 10.1177/2516600x20974121.

[4] T. Terttiaavini and T. S. Saputra, “Pendampingan Manajemen Pengelolaan Pasar Sekojo Dengan Membangun Market Management One Gate System (MMOGS),” *Selaparang J. Pengabdi. Masy. Berkemajuan*, vol. 7, no. 4, pp. 2543–2551, 2023, [Online]. Available: https://journal.ummat.ac.id/index.php/jpmb/article/view/19594/8211

[5] D. Marcelina, A. Kurnia, and T. Terttiaavini, “Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 293–301, 2023, doi: 10.57152/malcom.v3i2.952.

[6] T. Terttiaavini, Y. Hartono, E. Ermatita, and D. P. Rini, “Comparison of Simple Additive Weighting Method and Weighted Performance Indicator Method for Lecturer Performance Assessment,” *Mod. Educ. Comput. Sci.*, vol. 15, no. 2, pp. 1–11, 2023, doi: 10.5815/ijmecs.2023.02.01.

[7] Y. Kustiyahningsih *et al.*, “Integration K-Means clustering and AHP for recommendations batik MSMEs,” *E3S Web Conf.*, vol. 499, pp. 1–7, 2024, doi: 10.1051/e3sconf/202449901006.

[8] M. S. Kaiser, J. Xie, and V. S. Rathore, *Information and Communication Strategies for Competitive Technology (ICTCS 2021)*, Lecture No., vol. 401, no. Ictcs. Springer, 2023. doi: 10.1007/978-981-19-0098-3_14.

[9] Terttiaavini and T. S. Saputra, “Analisa Pelatihan Strategi Manajemen Penjualan Produk Umkm Menggunakan Digital Marketing Bagi Masyarakat Terdampak Covid-19 Di Kampung Keluarga Berhasil (Kb) Layang-Layang Palembang,” in *Seminar Nasional AVoER XII 2020*, Palembang: Fakultas Teknik Universitas Sriwijaya, 2020, pp. 18–19. [Online]. Available: http://ejournal.ft.unsri.ac.id/index.php/avoer/article/view/251

[10] T. Terttiaavini and T. S. Saputra, “Analisa Pelatihan Strategi Manajemen Penjualan Produk UMKM Menggunakan Digital Marketing Bagi Masyarakat Terdampak Covid-19 Di Kampung Keluarga Berhasil (Kb),” in *Seminar Nasional AVoER XII 2020*, 2020, pp. 18–19.

[11] T. Terttiavini, L. Hertati, Y. Yulius, and T. S. Saputra, “Strategi Digital Marketing dan Inovasi produk untuk meningkatkan daya saing UMKM Ikan Pedo Serbuk di Kabupaten Muratara,” Palembang, 2023. [Online]. Available: https://drive.google.com/file/d/1KlK9ef0ZALIe65CWnqi49cMNBIEF8wFT/view?usp=sharing

[12] S. Rochayatun, . S., and R. Bidin, “Mode of Entry Strategy on SMEs Internationalization in East Java: A Review of Literature,” *Asian J. Econ. Bus. Account.*, vol. 22, no. 15, pp. 20–32, 2022, doi: 10.9734/ajeba/2022/v22i1530626.

[13] T. Terttiaavini *et al.*, “Clustering Analysis of Premier Research Fields,” *Int. J. Eng. Technol.*, vol. 7, no. 4.44, p. 43, 2018, doi: 10.14419/ijet.v7i4.44.26860.

[14] K. P. Sinaga and M. Yang, “Unsupervised K-Means Clustering Algorithm,” *IEEE Access*, vol. 8, no. May, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796 Unsupervised.

[15] R. Suwanda, Z. Syahputra, and E. M. Zamzami, “Analysis of Euclidean Distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K,” *J. Phys. Conf. Ser.*, vol. 1566, no. 1, 2020, doi: 10.1088/1742-6596/1566/1/012058.

[16] F. Liu and Y. Deng, “Determine the Number of Unknown Targets in Open World Based on Elbow Method,” *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 5, pp. 986–995, 2021, doi: 10.1109/TFUZZ.2020.2966182.

[17] Terttiaavini, Y. Hartono, Ermatita, and D. P. Rini, “Comparison of Simple Additive Weighting Method and Weighted Performance Indicator Method for Lecturer Performance Assessment,” *Int. J. Mod. Educ. Comput. Sci.*, vol. 15, no. 2, pp. 1–11, 2023, doi: 10.5815/ijmecs.2023.02.01.

[18] L. Kovács and H. Ghous, “Efficiency comparison of Python and RapidMiner,” *Multidiszcip. Tudományok*, vol. 10, no. 3, pp. 212–220, 2020, doi: 10.35925/j.multi.2020.3.26.

[19] L. Magdalena and R. Fahrudin, “Penerapan Data Mining Untuk Koperasi Se-Jawa Barat Menggunakan Metode Clustering pada Kementerian Koperasi dan UKM,” *J. Digit*, vol. 9, no. 2, p. 190, 2020, doi: 10.51920/jd.v9i2.120.

[20] D. Marcelina, A. Kurnia, and T. Terttiaavini, “Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. October, pp. 293–301, 2023.

[21] S. A. Mustaniroh, L. S. Jauhari, and J. M. Maligan, “Strategi Pengembangan Klaster Ukm Produksi Bandeng Asap dengan Menggunakan Metode K-Means Clustering dan Fuzzy Ahp,” *J. Keteknikan Pertan. Trop. dan Biosist.*, vol. 8, no. 1, pp. 101–106, 2020, doi: 10.21776/ub.jkptb.2020.008.01.10.

[22] P. Dauni, P. Pratiwi, and R. T. Prasetio, “Pemetaan Keberlangsungan Hidup Umkm Guna Optimalisasi Bantuan Kredit Menggunakan Algoritma Fuzzy C-Means,” *J. Responsif Ris. Sains dan Inform.*, vol. 5, no. 1, pp. 61–69, 2023, doi: 10.51977/jti.v5i1.1051.

[23] D. Anton, T. Avini, A. Heryati, and H. Saputra, *Business Process Reengineering*, 1st ed. Tasikmalaya: Perkumpulan Rumah Cemerlang Indonesia, 2023.

[24] Hartatik *et al.*, *Data Science For Business (Pengantar & Penerapan berbagai Sektor)*, no. September 2016. 2023. [Online]. Available: https://www.data-science.ruhr/about_us/