International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 16, No.1 (Aug-24)

http://dx.doi.org/10.12785/ijcds/160165

Halal Supply Chain Risk using Unsupervised Learning Methods for Clustering Leather Industries

Rahmad Kurniawan[1], Fitra Lestari[2], Mawardi[3], Tengku Nurainun[2], Abu Bakar Abdul Hamid[4] and Tisha Melia[1]

[1]*Department of Computer Science, Universitas Riau, Pekanbaru, Indonesia*
[2]*Department of Industrial Engineering, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia*
[3]*Faculty of Sharia and Law, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia*
[4]*Putra Business School, Universiti Putra Malaysia, Serdang, Malaysia*

Received 5 Dec. 2023, Revised 4 May 2024, Accepted 7 May 2024, Published 10 Aug. 2024

Abstract: Cowhide plays a significant role in Indonesia's culinary and leather industries. It caters to the preferences of a predominantly Muslim population that emphasizes halal products. Regulatory authorities must understand its characteristics comprehensively to provide effective halal assurance to the diverse entities within Indonesia's leather industry. Traditional statistical methods for assessing halal compliance are inefficient due to the complexity and diversity of the leather industry's supply chain. This study addresses these challenges by employing unsupervised learning methods, specifically K-Means and Hierarchical clustering algorithms to analyze a dataset comprising 100 Cowhide Small and Medium Enterprises (SMEs) located in Garut Regency, West Java Province. This dataset includes 62 features that facilitate the clustering of these industries based on various halal risk factors. Experimental results indicate that the optimal number of clusters is four. The K-Means algorithm outperforms the Hierarchical clustering algorithm with a higher average silhouette score of 0.59 compared to 0.31. Furthermore, the K-Means algorithm demonstrates stability in clustering the data, making it a robust choice for this analysis. These clustering outcomes offer valuable insights into the SMEs operational characteristics and halal compliance risks, significantly enhancing the ability of regulatory authorities to implement effective halal assurance measures. Consequently, this study provides a robust framework for improving halal certification processes and aiding risk management within Indonesia's leather industry.

Keywords: K-means clustering, Hierarchical clustering, Cowhide, Leather Industries, SMEs, Halal

1. INTRODUCTION

Halal businesses have gained widespread popularity across various sectors, ranging from food and beverages to pharmaceuticals, cosmetics, tourism, finance, and even fashion [1]. In Indonesia, leather, especially cowhide, serves a dual purpose as a food source and a raw material for leather products [2]. These two categories of products are highly susceptible to halal-related issues, especially for companies that produce both types of products within a single production unit. The top layer of cowhide is utilized for craft products, while the inner layer is employed for cracker products. While the leather industry contributes substantially to the economy and experiences rapid growth, it also gives rise to intricate and severe risks within its supply chain [3].

Given the importance of ensuring that cowhide-based food products are free from non-halal materials such as pigskin or skin from other proscribed by Islamic law, producers of cowhide-based crafts must adhere to strict halal material guidelines. The persistent circulation of cowhide-based products made from pigskin in Indonesia poses a significant threat to the sustainability of these products. Consequently, it is important to verify the halal status of these products, including those within the cowhide industry, by implementing a robust halal assurance system. The Indonesian government has expressed its concern by enacting Law No. 33 of 2014, which mandates halal product assurance for all products distributed in Indonesia, both food and non-food.

Based on prior investigations, the leather industry encompasses a multifaceted business process that begins with slaughtering animals and is followed by separating hides. It culminates in the tanning process during leather manufacturing and ultimately produces leather sheets, as illustrated

E-mail address: rahmadkurniawan@lecturer.unri.ac.id, fitra.lestari@uin-suska.ac.id, mawardi@uin-suska.ac.id

in Figure 1. The Small and Medium Enterprises (SMEs) specializing in leather crafting process the leather to create various leather goods that are eventually distributed to consumers. The involvement of these business players (SMEs) across the supply chain represents a strategic approach to supply chain management. Elevating product quality standards within each SMEs is achievable by identifying risks, particularly halal-related risks, within these enterprises. One critical risk to address is the halal risk, ensuring consumers feel secure and comfortable using halal leather goods. Therefore, the identification of halal supply chain risks is essential for SMEs.

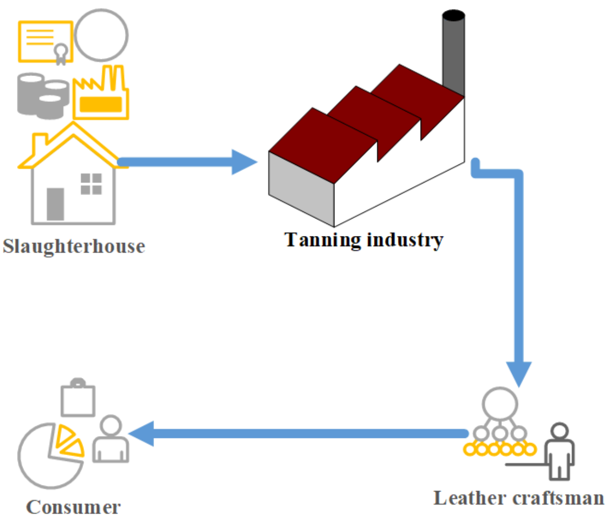

Figure 1. The previous investigation on the ideal supply chain within the leather industry

The assessment of halal compliance for leather craft products differs from that for cowhide-based food products. According to [4], the origin of the leather and the tanning process are the two primary factors determining the halal status of leather crafts. These factors are closely related to the significant halal integrity risks identified by [5], which encompass the status of raw materials, processing methods, wholesomeness, and shared facilities for halal and non-halal products. Previous study has predominantly focused on identifying and managing halal risks in various contexts, such as the red meat industry [6] or the frozen food industry [7].

The conventional methods of halal risk management, as delineated by [6] and further extended by [7], involve robust qualitative and quantitative analyses that focus on identifying and controlling risks at various stages of supply chains. The former study focuses on the red meat industry, advocating for the creation of detailed manuals and strict transportation policies to ensure the integrity of halal meat from Australia to Indonesia. The latter study provides insights into the frozen food sector, specifically milkfish brain products, proposing a framework that includes training on good manufacturing practices and developing halal standard operating procedures to mitigate risks.

A study emphasizes the importance of risk management in faith-based supply chains, particularly the halal supply chain [8]. The study identifies and prioritizes risk elements for managing the Halal Supply Chain (HSC) using a fuzzy best–worst method. The results indicate that production-related risks are highly significant.

While these strategies lay a strong foundation for ensuring halal compliance, the challenges posed by the increasing complexity and volume of data in global supply chains necessitate more sophisticated, scalable, and automated solutions. Machine learning offers a substantial advancement over traditional methods by providing a sophisticated analytical framework capable of processing large-scale data efficiently. This study builds on the foundational work of [6], [7] by utilizing unsupervised learning techniques, such as clustering algorithms, to identify and predict potential halal integrity risks through advanced pattern recognition.

The application of machine learning algorithms enables the proactive identification of risk factors in the supply chain, which traditional methods may overlook. For instance, machine learning can analyze historical data to forecast potential risk scenarios before they occur, allowing for preemptive mitigation strategies that are data-driven and tailored to emergent trends. Additionally, using these algorithms enhances the scalability of risk management processes, accommodating the expansive and evolving nature of international supply chains.

In essence, while the models proposed by [6], [7] provide valuable references for halal meat and frozen food auditing and procurement policies, the integration of machine learning in this research extends the capabilities of risk analysis. This approach offers a more dynamic, predictive, and comprehensive method for maintaining halal compliance in the global supply chains of various food products. By leveraging machine learning, this study increases the efficacy of halal assurance systems and equips stakeholders with a powerful tool for continuous improvement and adaptation in their risk management strategies.

Traditional statistical risk assessment has proven insufficient for assessing halal compliance risks in the diverse leather SMEs of Indonesia, particularly because of their inefficiency in handling extensive data [9]. This study addresses these limitations by employing unsupervised machine learning algorithms to cluster these industries based on comprehensive questionnaire data. By using clustering techniques, we aim to reveal underlying similarities and risk profiles among SMEs, enabling authorities and stakeholders to identify high-risk entities and tailor halal assurance measures accordingly effectively. This approach not only enhances the precision of risk management but also optimizes the allocation of resources to ensure sector-wide halal compliance. Furthermore, the clustering of leather SMEs serves as a method to gain insights into the characteristics of this sector, with clustered industries deemed to share similar characteristics. Consequently, competent authorities can provide more efficient halal assistance by identifying industries that do not adhere to halal principles in their production.

The study employs unsupervised learning approaches, i.e., K-means and Hierarchical clustering algorithms to cluster leather industries systematically based on identified halal risk factors. In order to assess the consistency within these clusters, the Silhouette score criterion is employed, which evaluates the degree of separation between clusters. By leveraging advanced data analysis techniques and this specific metric, we intent to uncover underlying patterns and correlations that may not be apparent through traditional risk assessment methods. This innovative approach is intended to provide a more nuanced understanding of the halal compliance challenges within the leather industry, ultimately facilitating the development of targeted and effective mitigation strategies.

2. RELATED WORKS

Machine learning has witnessed significant utilization across various domains for analyzing extensive datasets at risk [9]. Informatics and statistical analysis through machine learning have become popular approaches for solving specific trained problems. Studies have employed machine learning techniques to address halal-related issues as per the existing literature. For instance, two studies have conducted sentiment analysis of halal products using Twitter data [10], [11], while deep learning methods have been employed to detect non-halal ingredients in food [12]. However, these studies primarily utilize supervised learning methods. Supervised learning, although effective, poses challenges when applied to unlabeled data, such as data from questionnaires in the cowhide research industry in the Garut Regency. Conversely, unsupervised learning techniques are well-suited for handling unlabeled data.

Unsupervised learning methods have attracted significant attention for their ability to analyze questionnaire data, providing valuable insights into patterns, structures, and relationships within datasets. Research exemplified by studies [13], [14] has demonstrated the efficacy of unsupervised machine learning in revealing hidden clusters and heterogeneity within diverse contexts, including patient data and secure Internet of Things (IoT) environments, respectively.

One widely used unsupervised learning method is clustering, which can group unlabeled data effectively, irrespective of the number of attributes [15]. Among the array of data analysis techniques, clustering is a method for categorizing data into groups sharing similar characteristics [16]. Clustering algorithms represent a potent approach for extracting valuable insights by grouping data points.

A study presented innovative approaches to cyber risk assessment that incorporate attributes of the digital supply chain [17]. The study used unsupervised clustering techniques and deep reinforcement learning algorithms to identify high-risk software companies based on their relative position in the supply chain.

Both the K-means and Hierarchical Clustering algorithms have been employed to address problems across various domains, such as microarray efficiency [18], DNA sequence analysis [19], astronomy [20], and data visualization [21]. The versatility of clustering algorithms, exemplified by their application in diverse domains, underscores their efficacy as a potent tool for extracting valuable insights from complex datasets.

In a recent study, remarkable results have been achieved through clustering using the K-means algorithm to cluster 25 mammals [22]. The study also evaluated various approaches to determine the optimal number of clusters, concluding that the approach with the fewest clusters was the most appropriate. Conversely, the hierarchical clustering algorithm has been applied extensively in data analysis. This algorithm creates dendrograms based on data distances, facilitating data grouping. Even complex fields such as research astronomy leverage hierarchical algorithms, spanning a wide range of scales from asteroids and molecular clouds to galaxies and galaxy clusters.

A recent study used a hierarchical algorithm to support decision-makers in comprehending the distribution of educational variables across Yemen [23]. The algorithm aided in extracting new intrinsic educational insights. The advantages of the hierarchical algorithm are evident in its ability to illustrate data grouping using dendrograms. Similarly, unsupervised learning has been employed to analyze the risk of COVID-19 across more than 200 countries worldwide [9], indicating that machine learning algorithms can be leveraged to identify hidden data patterns. The hierarchical approach enables the grouping of data objects into a hierarchical structure resembling a tree. Each level or node within this hierarchy represents a distinct cluster [16].

Moreover, studies comparing K-means and hierarchical algorithms empirically are particularly intriguing. A hierarchical algorithm is often used as a baseline and compared against a given dataset. Hence, we have employed the efficient K-means and Hierarchical clustering algorithms to cluster the cowhide industry based on questionnaire results for our study.

Furthermore, we also focus on applying two widely used clustering algorithms, K-means and Hierarchical Clustering, to the context of the Cowhide SMEs Industry in Garut. Several studies underscore the critical importance of validating clustering results using the Silhouette score, as demonstrated in previous studies [24], [25], [26], [27].

Clustering algorithms have been extensively applied in various domains, but their performance can vary depending on the dataset and the choice of parameters. Therefore, it is imperative to validate the clustering results rigorously. One standard metric for assessing the quality of clusters is the Silhouette score, which measures the similarity of data points within clusters compared to close clusters [24] . The motivation for emphasizing the justification of clustering outcomes employing the Silhouette score is supported by several relevant studies. In a comparative study [24], the authors evaluated the performance of K-Means clustering against another technique, CLARA clustering, using the Silhouette score on the Iris dataset.

The findings highlighted the significance of Silhouette analysis as a cluster validation measure, shedding light on the effectiveness of different clustering methods [24], Similarly, in a study on load profiles clustering methods [25], the Silhouette score criterion was employed to assess the consistency within clusters generated by Density-Based Spatial Grouping of Purposes with Noise, Hierarchical cluster analysis, and K-means clustering. The average Silhouette scores were crucial in ranking the clustering methods based on their performance [25].

Moreover, clustering techniques have been enhanced and adapted to address specific challenges. For instance, a density clustering algorithm based on the Silhouette coefficient was proposed [26] to improve the accuracy of edge point division in the DBSCAN algorithm. This innovation demonstrates the Silhouette coefficient's significance as a criterion for refining clustering results, especially in scenarios where traditional methods face limitations [24] .

Additionally, the Silhouette index has been explored in conjunction with the K-Harmonic Means method to group remote sensing datasets effectively [27]. This approach showcases the Silhouette index's ability to determine the correct number of clusters in scenarios with varying degrees of cluster overlap [27]. A study highlights the challenges of identifying compact and well-separated clusters within datasets, a task many clustering algorithms grapple with [28]. While traditional clustering approaches focus on optimizing clustering objective functions that capture intra-cluster similarity and inter-cluster dissimilarity, these functions alone may not guarantee the discovery of distinctly separated and compact clusters.

Researchers have increasingly relied on cluster validity indices like the Silhouette score for their versatility and efficiency in evaluating clustering without needing a training set [28], [29], [30], [31], [32]. By leveraging this metric, we aim to provide a robust evaluation of the performance of K-means and Hierarchical Clustering algorithms in clustering the Cowhide SMEs Industry data in Garut, ultimately contributing to a more reliable and insightful analysis.

Furthermore, the current literature on halal supply chain management within the leather industry predominantly focuses on compliance and certification processes. However, there is a significant research gap in applying unsupervised learning algorithms for risk analysis tailored to the nuanced characteristics of the leather industry's halal supply chain. Addressing this gap is crucial for advancing effective halal assurance measures and enhancing the robustness of the industry's risk management strategies.

Therefore, this study proposes a novel approach using unsupervised learning algorithms to cluster the Leather SMEs industry in Garut, Indonesia, based on halal risk factors. This approach not only provides valuable insights into the industry's characteristics but also facilitates the efficient implementation of halal assistance, ultimately enhancing halal compliance within the cowhide industry. Additionally, we provide a robust evaluation using Silhouette scores of the performance of K-means and Hierarchical Clustering algorithms in clustering the Cowhide SMEs Industry data in Garut, ultimately contributing to a more reliable and insightful analysis.

3. MATERIAL AND METHODS

We specifically identified and included halal risk factors that are critical in assessing the halal integrity of the supply chain within the leather industry. These factors were incorporated into the dataset comprising 100 Cowhide Small and Medium Enterprises (SMEs) in Garut Regency, West Java Province, Indonesia. The questionnaire comprised three sections, each serving distinct purposes.

Section A, focusing on SMEs company profile, encompassed several items such as *ID number*, *company name*, *address*, *telephone number*, *email*, *industry type*, *number of employees*, *years in operation*, *sales and distribution*. This section was aimed at gathering foundational information about each SMEs. Information about sales and distribution channels was also collected to provide insights into each SMEs business operations and market reach.

Subsequently, Section B comprised 16 questionnaire items about implementing integrated quality and environmental management. This section delved into practices related to quality and environmental management. The focus was on understanding how each SMEs integrates these aspects into their operations, which are essential for assessing their compliance with international standards and environmental sustainability practices.

Section C contained 19 items related to halal implementation. It covered various processes and compliance measures that ensure the halal integrity of the leather produced, from raw material sourcing to final product handling.

Finally, Section D, which is dedicated to performance, consists of 17 items. This section aimed to assess the effectiveness of the implemented practices in terms of operational success and compliance with halal standards.

The chosen risk factors are integral for maintaining the halal integrity of leather products. By clustering the SMEs based on these factors, our study assesses how each enterprise complies with halal standards, thus identifying potential operational risks. This clustering enables stakeholders, including regulatory bodies and SMEs, to pinpoint improvement areas and ensure rigorous compliance with halal certification requirements.

The analysis utilized unsupervised learning methods, specifically K-means and hierarchical clustering. It leveraged these risk factors to segment the industries into distinct groups, revealing variations in risk levels and compliance. This differentiation assists in targeted interventions and enhances the overall management of halal assurance measures within the leather industry. Figure 2 illustrates the key steps involved in the methodology.

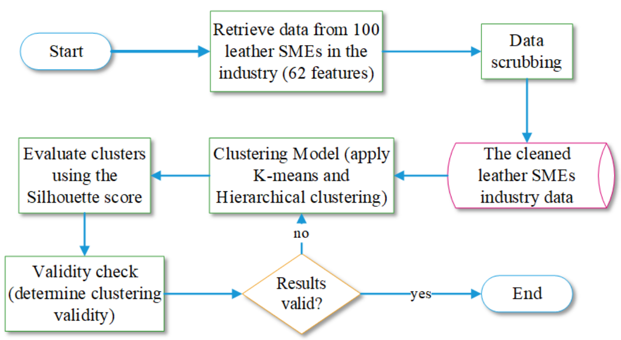

Figure 2. Methodology of an unsupervised learning algorithm for clustering the Cowhide SMEs Industry in Garut Regency, Indonesia

The methodology employed in this study is presented in Figure 2, which provides a brief overview of the unsupervised learning algorithm used to cluster the Cowhide SMEs Industry in Garut.

A. Data Scrubbing

We collected data from 100 Cowhide SMEs Industries in Garut. Data scrubbing techniques were applied during the modeling process to enhance data recognition and improve the unsupervised learning algorithm's learning process. The respondents obtained the dataset directly, ensuring all values were present. However, data transformation was necessary to obtain standardized data. Hence, we employed the *one-hot encoding* technique.

The *one-hot encoding* method transforms categorical data into numerical data from 0 and 1. Features with multiple parameters are transformed into new segments with data values between 0 and 1. The new features are assigned a value of 1 if they match the categorical data and 0 otherwise. This transformation process is crucial as unsupervised learning algorithms, such as clustering algorithm, operate exclusively on numeric features.

B. Clustering Model

In this study, we employed clustering techniques to categorize data into meaningful groups, using two predominant algorithms: K-means and hierarchical clustering. Both methodologies were implemented in *Python*, leveraging its robust libraries for machine learning and data analysis. The computational experiments were conducted on a machine equipped with Windows 10 Pro OS, a 64-bit system, an Intel Core i7 Gen3 processor with 8 GB RAM, a 250 GB SSD, and an Intel HD Graphics 4000 with 2 GB of video memory. This setup ensured the efficient processing of the data-intensive tasks inherent to machine learning algorithms.

The Leather SMEs Industry in Garut can be clustered using the K-means and hierarchical clustering algorithms. Let $\delta_s$be the number of Cowhide SMEs Industry in Garut data points in a cluster, and δ be the total number of cluster centers. The K-means clustering algorithm minimizes the squared error function $\|\beta_s - \gamma_t\|$, where β_s is a data point and γ_t is a cluster center, with the objective function$\alpha(\beta, \gamma, \delta)$ given by:

$$\alpha(\beta, \gamma, \delta) = \sum_{s=1}^{\delta} \sum_{t=1}^{\delta_s} (\|\beta_s - \gamma_t\|)^2 \tag{1}$$

The dataset comprises Cowhide SMEs *ID number*, *company name*, *address*, *telephone number*, *email, industry type*, *number of employees*, *years in operation*, *certifications*, *sales & distribution*, and 52 features from questionnaires. We tested various cluster sizes with $m = 2, 3, 4$. To begin, we randomly select cluster centers in step 1. In step 2, we estimate the distance between each Cowhide SMEs Industry in Garut data point, β_s, and cluster centers, γ_t. In step 3, we substitute β_s with the cluster center, γ_t, that has the lowest distance α compared to all the other cluster centers, δ. Next, we re-estimate the new cluster center with the average value of $\gamma_s = \frac{1}{\delta_s} \sum_{t=1}^{\delta_s} \beta_s$, and the distance between each Cowhide SMEs Industry in Garut data point, β_s, and the new cluster centers, γ_{new}, in steps 4 and 5. Finally, the algorithm will stop if no Cowhide SMEs Industry in the Garut data point is substituted or proceeds to step 3.

We utilized the *Scikit-learn* library for *K-means* clustering, a versatile tool in the Python ecosystem designed for efficiently implementing machine learning algorithms. The *K-means* class from *Scikit-learn* was instantiated with a predefined number of clusters, and the algorithm was executed to partition the dataset into distinct groups based on the minimization of within-cluster variances, also known as inertia. The initialization method and the number of iterations were set according to the dataset's characteristics to optimize convergence.

In the Hierarchical clustering algorithm we began the Hierarchical clustering algorithm by initializing the cluster $\beta = \{\beta_1, \cdots, \beta_n\}; n = 100$ be the set of Cowhide SMEs Industry in Garut data points containing of *Cowhide SMEs ID number*, *company name*, *address*, *telephone number*, *email, industry type*, *number of employees*, *years in operation*, *certifications*, *sales & distribution*, and 52 features from questionnaires, $\delta = \{\delta_1, \cdots, \delta_m\}$ be the set of center points. Next, we conducted an experiment of hierarchical clustering on 100 rows of cowhide SMEs data dataset, dividing it into $m = 4$ clusters using Ward's method. To calculate the dissimilarity between clusters, we computed the Euclidean distance between each pair of clusters using Formula (2):

$$\beta(\beta_i, \beta_n) = distance(\beta_i, \beta_n) = \text{for all } i, n \text{ such that } i < n \tag{2}$$

The hierarchical clustering was implemented using the 'scipy.cluster.hierarchy' module from the SciPy library, which provides functions for agglomerative clustering. We computed the distance matrix using the pdist function, which measures the Euclidean distances between data points.

In simple ways, we need to implement the following steps:

1) Initialization
   - Randomly select δ initial cluster centers for K-means, where δ represents the number of desired clusters.
   - Initialize the Hierarchical Clustering algorithm with individual data points as initial clusters.
2) K-means Clustering
   - Apply the K-means algorithm to partition the data into δ clusters.
   - Calculate the distance between each data point β_s and the cluster centers γ_t using the Euclidean distance.
   - Assign each data point β_s to the cluster center γ_t with the smallest distance.
   - Update the cluster centers γ_t based on the average value of the data points within each cluster.
   - Repeat the K-means steps until convergence is achieved.
3) Hierarchical Clustering
   - Calculate the dissimilarity (distance) between clusters using Euclidean distance.
   - Merge the closest pair of clusters based on the calculated dissimilarity.
   - Update the distance matrix to reflect the newly formed clusters.
   - Repeat the merging and distance matrix update steps until the desired number of clusters m is reached or a specific criterion is met.

   Finally, scatter plots were generated using the *matplotlib* to visually compare the outcomes of both clustering *algorithms.pyplot* module. These plots displayed the clusters formed by each algorithm, using different colors to represent various clusters. This visual representation facilitated an intuitive analysis of the clustering structures and their respective efficacies in segregating the dataset into coherent groups.
4) *Integration*
   Combine the K-means and Hierarchical Clustering results to obtain a final set of clusters. This integration can be achieved by mapping the clusters obtained from K-means to corresponding clusters in the hierarchical dendrogram.

C. *Testing and Evaluation*

Several parameters were carefully selected as experimental settings, including the distance function, initialization method, maximum iteration, total number of items, number of clusters, the within-cluster sum of squared errors, and time is taken. The silhouette method was applied to interpret and validate the consistency within the data clusters and measure the proximity of an object to its own group compared to other clusters [9]. The silhouette coefficient is calculated using the following formula 3 [33]:

$$\text{Silhouette Coefficient } = \frac{(b - a)}{max(a, b)} \tag{3}$$

In the formula, *(a)* represents the mean distance of a Cowhide SMEs Industry in Garut, West Java, to other Cowhide SMEs Industry data within the same cluster, while *(b)* represents the mean distance of the Cowhide SMEs Industry to the nearest instances in the next closest cluster.

4. Results and Discussion

We conducted three experiments in this section, as detailed in I. The initial experiment utilized the standard K-means algorithm to cluster the Cowhide SMEs Industry in Garut.

Based on the results presented in I, the best performance was achieved when the number of clusters m=4. Therefore, subsequent experiments for clustering the Cowhide SMEs Industry in Garut were conducted with number of clusters m=4 for both the K-means and Hierarchical clustering algorithms. As shown in Table I, it can be observed that a smaller number of clusters yielded higher silhouette scores, which are closer to 1.

Evaluation based on the silhouette score parameter revealed that the K-means algorithm outperformed the hierarchical algorithm, with average silhouette scores of 0.59 and 0.31, respectively. Figures 3 and 4 illustrate the silhouette scores of the 100 Cowhide SMEs Industry entities in Garut, West Java Province, Indonesia, clustered using the K-means and Hierarchical clustering algorithms when m=4.

Furthermore, Figures 5(a) and 5(b) depict the consistency within the data clusters when m=4, utilizing both the K-Means and Hierarchical algorithm clustering for the 100 Cowhide SMEs Industry entities in Garut.

Based on Figure 5(a), it is evident that unsupervised learning has successfully grouped the Cowhide SMEs Industry entities in Garut based on the data. Four clusters were formed, with cluster 1 (C1=10) exhibiting the highest level of consistency, while cluster 3 (C3=24) appeared to be visually less consistent than the other clusters. In Figure 5(b), cluster 1 (C3=10) displayed the highest consistency, while cluster 3 (C1=56) appeared to be visually less consistent compared to the other clusters.

Table II presents the generated silhouette scores to ana-

TABLE I. EXPERIMENT SETTINGS OF THE K-MEANS CLUSTERING ALGORITHM

Experiment settings	Run 1	Run 2	Run 3
Distance function	Euclidean distance	Euclidean distance	Euclidean distance
Initialization method	Random	Random	Random
Max iteration	300	300	300
Total number of items (Cowhide SMEs Industry in Garut)	100	100	100
Number of clusters (m)	2	3	4
Within cluster sum of squared errors (α)	753.4	722.4	661.5
Average of silhouette	0.57	0.58	0.59
Time taken	0.06 seconds	0.01 seconds	0.02 seconds

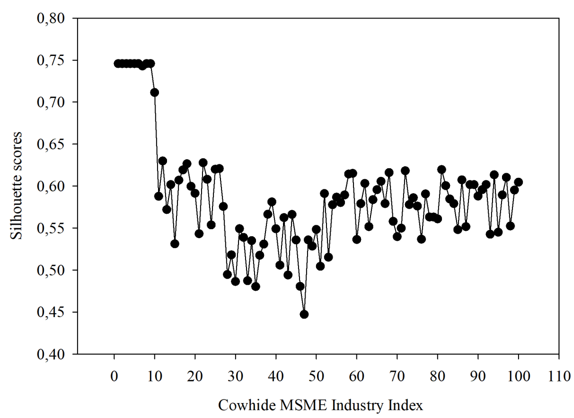

Figure 3. Silhouette scores when m=4 for the 100 Cowhide SMEs Industries in Garut using the K-Means clustering algorithm

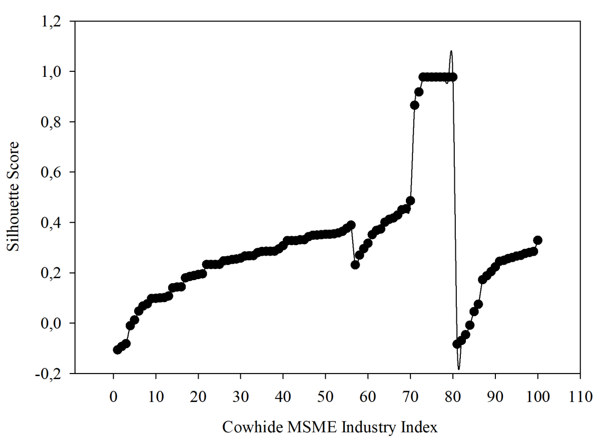

Figure 4. Silhouette scores when m=4 for the 100 Cowhide SMEs Industries in Garut using the Hierarchical clustering algorithm

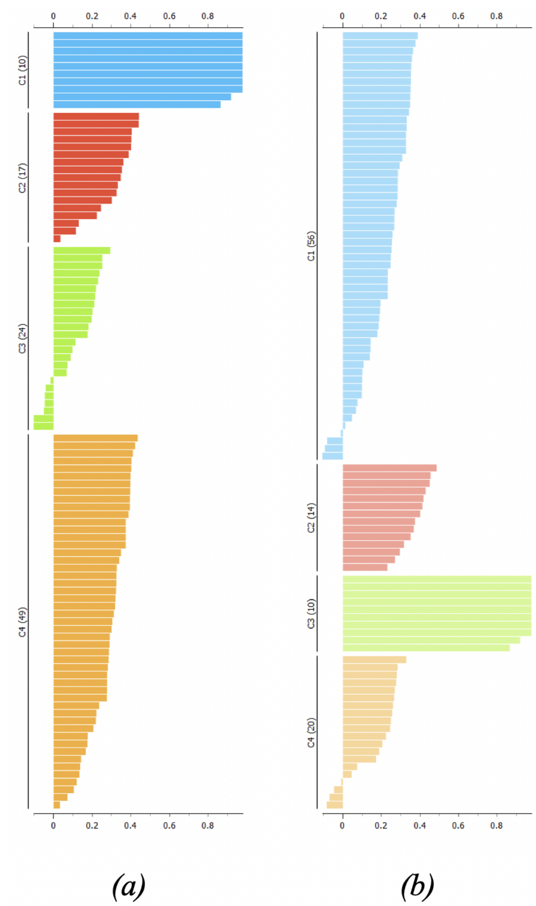

Figure 5. Consistency within clusters of (a) the K-Means clustering algorithm and (b) the Hierarchical clustering algorithm when m=4 for the 100 Cowhide SMEs Industries in Garut with the Euclidean distance metric

lyze the consistent cluster results further, as shown in Figure 5. These scores indicate the proximity of each Cowhide SMEs Industry entity to its respective cluster compared to other clusters. Silhouette scores close to 1 indicate that the data instances are located near the cluster's center, while scores close to 0 suggest instances located on the borders between two clusters.

In the K-Means algorithm, cluster C1 demonstrates the highest level of consistency, while in the Hierarchical algorithm, cluster C3 exhibits the highest consistency. Interest-

TABLE II. K-Means Algorithm Results for Cluster 1 (C1) of Leather SMEs Industry When the Number of Clusters is Set to m=4

No.	Cowhide SMEs Industry	Clusters	Silhouette Scores
1.	Ba S	C1	0.745795021
2.	HW S	C1	0.745795021
3.	Gu S	C1	0.745795021
4.	Ha S	C1	0.745795021
5.	In S	C1	0.745795021
6.	Pa S	C1	0.742960487
7.	It S	C1	0.745795021
8.	Pu S	C1	0.745795021
9.	Ra S	C1	0.745795021
10.	Si S	C1	0.711348127

ingly, despite using different algorithms, both yield similar clustering outcomes for the Cowhide SMEs Industry in Garut, West Java Province, Indonesia. Table III presents the results of the Hierarchical clustering algorithm for Cluster 3 (C3).

Based on the findings from Tables II and III, the clustering of the Cowhide SMEs Industry in Garut, West Java Province, Indonesia, is acceptable, as the data instances are closely located in the centers of their respective clusters. Subsequently, an experiment was conducted to organize the data of the SMEs Industry.

Considering the clustering results obtained from both the K-means and Hierarchical clustering algorithms, it can be concluded that both algorithms deliver satisfactory clustering outcomes, each with advantages. As depicted in Table III, the Hierarchical clustering algorithm yields generally higher silhouette scores for Cluster 3 (C3) than the K-means algorithm. However, the K-means algorithm, as a whole, demonstrates more consistent cluster results across the four clusters, as evident in the visual comparisons in Figures 5(a), 5(b), 6, and 7.

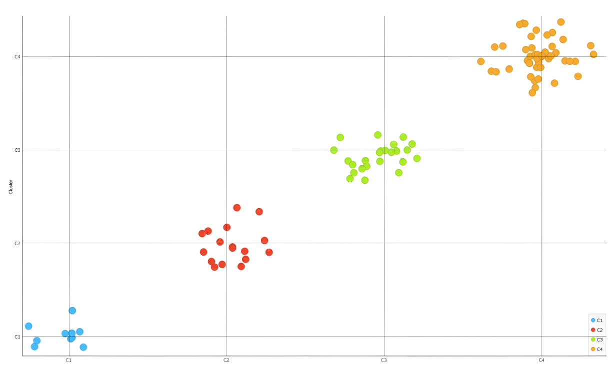

Figure 6. Visualization of clustering results using the K-means algorithm with the number of clusters set to m=4

Then, we experimented on Hierarchical clustering using box plot visualization. Figure 8 shows the box plot results to aid in the assessment of the algorithm's success in segregating data into clusters with distinct characteristics.

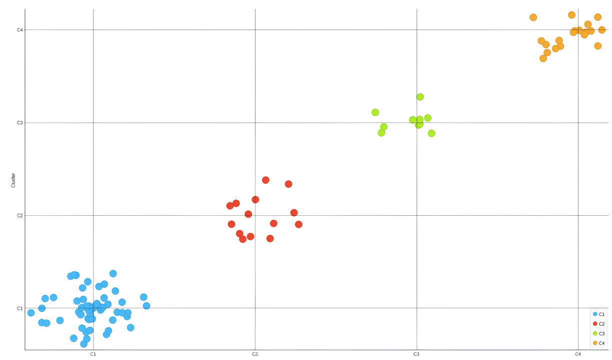

Figure 7. Visualization of clustering results using the Hierarchical clustering algorithm with the number of clusters m=4

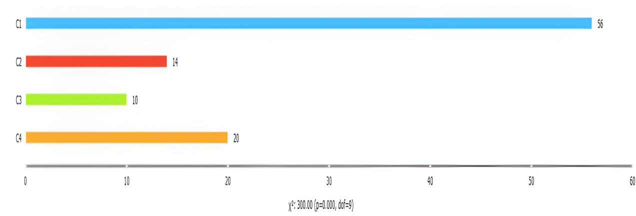

Figure 8. Visualization of box plot in the Hierarchical Clustering algorithm

According to Figure 8, the results of the Hierarchical Clustering analysis reveal the formation of distinct clusters labelled as C1, C2, C3, and C4. Each cluster is characterized by a specific number of data points, with C1 comprising 56 data points, C2 containing 14 data points, and both C3 and C4 each encompassing 20 data points.

Furthermore, the statistical information is represented as X^2 : 300.00 ($p = 0.000, dof = 9$), indicating the significance of the clustering results. The X^2 value denotes a computed statistical metric and suggests the effectiveness of the clustering algorithm. The p-value of 0.000 signifies a highly significant difference between the clusters, while the degree of freedom (dof) value of 9 reveals the complexity

TABLE III. Hierarchical Clustering Algorithm Results for Cluster 3 (C3) of Leather SMEs Industry When the Number of Clusters is Set to m=4

No.	Cowhide SMEs Industry	Clusters	Silhouette Scores
1.	Ba S	C3	0.97737016
2.	HW S	C3	0.97737016
3.	Gu S	C3	0.97737016
4.	Ha S	C3	0.97737016
5.	In S	C3	0.97737016
6.	Pa S	C3	0.918329608
7.	It S	C3	0.97737016
8.	Pu S	C3	0.97737016
9.	Ra S	C3	0.97737016
10.	Si S	C3	0.865374887

of the statistical analysis involved.

These results are pivotal for understanding the clustering algorithm's output quality. However, it is worth noting that despite both algorithms yielding satisfactory results in clustering the leather industry data, the k-means clustering algorithm demonstrates superior performance in terms of silhouette score when compared to the hierarchical clustering algorithm. Therefore, for a more comprehensive interpretation of the results, it is essential to emphasize the findings derived from the k-means clustering algorithm.

Based on the results obtained from the K-means algorithm, Cluster 1 (C1) comprises industries primarily involved in food production. Conversely, the craft industry dominates the second and third clusters. The final cluster encompasses industries of various types, with the leather tanning industry being grouped within this cluster. Considering the halal risks within the cowhide industry, addressing them specifically regarding raw materials, processing, transportation, and facilities is essential. The materials must originate from cows slaughtered following Islamic practices for cowhide-based food production.

In the context of Halal management, it is essential to identify high-risk industries that require special attention for halal assistance. The clustering algorithm can be utilized to determine industry clusters. When the industry data are processed automatically, they form their respective clusters. For example, the Cowhide SME Industry data with the closest distances to each other will be assigned to the same cluster, as demonstrated in Figure 9.

As per Fatwa No. 56 issued by the Indonesian Ulama Council (MUI) in 2014, halal leather products must not only originate from permissible animals and undergo a halal slaughtering process but also involve a tanning procedure devoid of non-halal substances. Considering that the tanning process employs various additional materials or chemicals that might contain non-halal substances, prioritizing the resolution of this issue within the C4 cluster becomes essential.

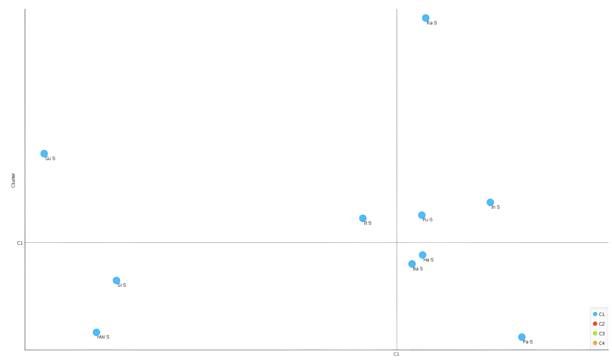

Figure 9. Assignment of multiple Cowhide SMEs Industries to the first cluster (C1) by the K-means algorithm when the number of clusters is set to m=4

Certain craft producers customize products based on orders, so the risk of contamination throughout the processing, shipping, and warehousing stages is higher than in cowhide-based food production. Consequently, this category of halal risks is of more significant concern for other clusters, notably C2, C3, and C4, as highlighted by the K-means algorithm results. Moreover, these clusters should be educated about the cleaning procedures (especially tanning) for shared facilities utilized in halal and non-halal leather-based production.

The involvement of 100 SMEs in the halal leather industry necessitates the identification of their halal risks. Thus, clustering SMEs is imperative to streamline policymaking to enhance the quality of halal standards. This study categorizes SMEs into clusters. Identifying these clusters will simplify policymaking for SMEs and decision-makers, facilitating their efforts to support global halal standardization.

5. CONCLUSION

This study has effectively utilized K-means and Hierarchical clustering algorithms to analyze halal risk factors in the Cowhide SMEs industry in Garut, West Java, Indonesia. By clustering industry data, this research offers profound insights into halal compliance, addressing critical concerns for Indonesia's substantial Muslim population and its stringent requirements for halal products. A significant finding from this study is the identification of the tanning process as a pivotal element in maintaining halal integrity, in line with the Indonesian Ulama Council's Fatwa No. 56 of 2014 regarding permissible materials and procedures.

The results provide a robust strategic framework for halal management and regulatory authorities to enhance halal compliance and risk assessments across similar industries. The deployment of unsupervised learning algorithms has enabled a deeper understanding of industry-specific risks and supported the development of more effective halal certification processes.

This research provides critical insights that are beneficial to various stakeholders in the leather industry. Regulatory authorities can refine their oversight and compliance standards more effectively, allocating resources to areas with higher risks. SMEs can use these insights to bolster their halal compliance practices and operational efficiencies, thereby enhancing their market access and competitive edge. For consumers, this study boosts confidence in the integrity of halal-certified leather products, facilitating more informed purchasing decisions aligned with their religious and ethical values. Overall, this application of data-driven analysis not only advances halal compliance within the leather industry but also sets a precedent for other sectors aiming to optimize regulatory and compliance practices through innovative technological approaches.

Moreover, this study highlights the critical role of the halal supply chain in guiding policymakers and SMEs toward standardization and improved productivity. Looking forward, expanding the dataset, exploring real-time monitoring systems, and fostering engagements with industry stakeholders are recommended to enhance future research. Such collaborative efforts could lead to the development of proactive risk mitigation strategies, further bolstering the growth and sustainability of the halal industry.

6. ACKNOWLEDGMENT

Directorate of Islamic Higher Education with Research Grant No. 6008/2022, Universitas Islam Negeri Sultan Syarif Kasim Riau, Center of Islamic Data Science and Continues Improvement (CIDSCI) and Computer Science Department of Universitas Riau support this work.

References

[1] M. S. Hossain, M. F. Rahman, M. K. Uddin, and M. K. Hossain, "Customer sentiment analysis and prediction of halal restaurants using machine learning approaches," *Journal of Islamic Marketing*, no. ahead-of-print, 2022.

[2] D. Amertaningtyas, "Mini review : Pengolahan kerupuk "rambak" kulit di indonesia," *Jurnal Ilmu-Ilmu Peternakan Universitas Brawijaya*, vol. 21, no. 3, pp. 18–29, 2011.

[3] A. Raihan, F. Islam, and S. M. Ali, "Modelling of supply chain risks in the leather industry," in *Proceedings of the 2019 IEOM International Conference, Toronto, Canada, October*, 2019, pp. 23–25.

[4] I. Jaswir, E. A. Rahayu, N. Yuliana, and A. Roswiem, "Daftar referensi bahan-bahan yang memiliki titik kritis halal dan substitusi bahan non-halal," *Jakarta: Komite Nasional Ekonomi dan Keuangan Syariah*, 2020.

[5] S. Khan, A. Haleem, and M. I. Khan, "Risk assessment model for halal supply chain using an integrated approach of ifn and d number," *Arab Gulf Journal of Scientific Research*, vol. 41, no. 3, pp. 338–358, 2022.

[6] U. Maman, A. Mahbubi, and F. Jie, "Halal risk mitigation in the australian–indonesian red meat supply chain," *Journal of Islamic Marketing*, vol. 9, no. 1, pp. 60–79, 2018.

[7] D. Kristanto and D. A. Kurniawati, "Development of halal supply chain risk management framework for frozen food industries," *Journal of Islamic Marketing*, vol. 14, no. 12, pp. 3033–3052, 2023.

[8] A. Sarwar, A. Zafar, and A. Qadir, "Analysis and prioritization of risk factors in the management of halal supply chain management," *Discover Sustainability*, vol. 2, no. 1, p. 30, 2021.

[9] R. Kurniawan, S. N. H. S. Abdullah, F. Lestari, M. Z. A. Nazri, A. Mujahidin, and N. Adnan, "Clustering and correlation methods for predicting coronavirus covid-19 risk analysis in pandemic countries," in *2020 8th International Conference on Cyber and IT Service Management (CITSM)*. IEEE, 2020, pp. 1–5.

[10] A. Feizollah, S. Ainin, N. B. Anuar, N. A. B. Abdullah, and M. Hazim, "Halal products on twitter: Data extraction and sentiment analysis using stack of deep learning algorithms," *IEEE Access*, vol. 7, pp. 83 354–83 362, 2019.

[11] R. Setik, R. M. T. R. L. Ahmad, and S. Marjudi, "Exploring classification for sentiment analysis from halal based tweets," in *2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*. IEEE, 2021, pp. 1–6.

[12] H. Fadhilah, E. C. Djamal, R. Ilyas, and A. Najmurrokhman, "Non-halal ingredients detection of food packaging image using convolutional neural networks," in *2018 International symposium on advanced intelligent informatics (SAIN)*. IEEE, 2018, pp. 131–136.

[13] M. Abdel-Basset, N. Moustafa, H. Hawash, W. Ding, M. Abdel-Basset, N. Moustafa, H. Hawash, and W. Ding, "Supervised deep learning for secure internet of things," *Deep Learning Techniques for IoT Security and Privacy*, pp. 131–166, 2022.

[14] S. Ferro, D. Bottigliengo, D. Gregori, A. S. Fabricio, M. Gion, and I. Baldi, "Phenomapping of patients with primary breast cancer using machine learning-based unsupervised cluster analysis," *Journal of personalized medicine*, vol. 11, no. 4, p. 272, 2021.

[15] R. Baruri, A. Ghosh, R. Banerjee, A. Das, A. Mandal, and T. Halder, "An empirical evaluation of k-means clustering technique and comparison," in *2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon)*. IEEE, 2019, pp. 470–475.

[16] P. J. Kaur *et al.*, "Cluster quality based performance evaluation of hierarchical clustering method," in *2015 1st International conference on next generation computing technologies (NGCT)*. IEEE, 2015, pp. 649–653.

[17] B. M. Siegel, "Innovative supply chain cyber risk analytics: Unsupervised clustering and reinforcement learning approaches," Ph.D. dissertation, Massachusetts Institute of Technology, 2023.

[18] T.-S. Chen, T.-H. Tsai, Y.-T. Chen, C.-C. Lin, R.-C. Chen, S.-Y. Li, and H.-Y. Chen, "A combined k-means and hierarchical clustering method for improving the clustering efficiency of microarray," in *2005 International symposium on intelligent signal processing and communication systems*. IEEE, 2005, pp. 405–408.

[19] L. Muflikhah, W. F. Mahmudy *et al.*, "Dna sequence of hepatitis b virus clustering using hierarchical k-means algorithm," in *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*. IEEE, 2019, pp. 1–4.

[20] H. Yu and X. Hou, "Hierarchical clustering in astronomy," *Astronomy and Computing*, vol. 41, p. 100662, 2022.

[21] L. M. Cabezas, R. Izbicki, and R. B. Stern, "Hierarchical clustering: Visualization, feature importance and model selection," *Applied Soft Computing*, vol. 141, p. 110303, 2023.

[22] Matt, "10 tips for choosing the optimal number of clusters," Towards Data Science, 2019, accessed Apr. 16, 2020. [Online]. Available: 10-tips-for-choosing-the-optimal-number-of-clusters-277e93d72d92

[23] F. Alqasemi, A.-H. Salah, A. Aqlan, K. M. Alalayah, Z. Almotwakl, and M. Hadwan, "Education data mining for yemen regions based on hierarchical clustering analysis," in *2021 International Conference of Technology, Science and Administration (ICTSA)*. IEEE, 2021, pp. 1–4.

[24] T. Gupta and S. P. Panda, "Clustering validation of clara and k-means using silhouette & dunn measures on iris dataset," in *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*. IEEE, 2019, pp. 10–13.

[25] A. Bosisio, A. Berizzi, A. Morotti, B. Greco, G. Iannarelli, C. Moscatiello, C. Boccaletti, and H. Noriega, "Performance assessment of load profiles clustering methods based on silhouette analysis," in *2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*. IEEE, 2021, pp. 1–6.

[26] G. Jin-Heng, L. Jia-Xiang, Z. Zhen-Chang, and L. Han-Yu, "Cdb-scan: Density clustering based on silhouette coefficient constraints," in *2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI)*. IEEE, 2022, pp. 600–605.

[27] H. Mahi, N. Farhi, K. Labed, and D. Benhamed, "The silhouette index and the k-harmonic means algorithm for multispectral satellite images clustering," in *2018 international conference on applied smart systems (ICASS)*. IEEE, 2018, pp. 1–6.

[28] A. M. Bagirov, R. M. Aliguliyev, and N. Sultanova, "Finding compact and well-separated clusters: Clustering using silhouette coefficients," *Pattern Recognition*, vol. 135, p. 109144, 2023.

[29] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, p. 759, 2021.

[30] L. Lenssen and E. Schubert, "Medoid silhouette clustering with automatic cluster number selection," *Information Systems*, vol. 120, p. 102290, 2024.

[31] M. B. Slimene and M.-S. Ouali, "Anomaly detection method of aircraft system using multivariate time series clustering and classification techniques," *IFAC-PapersOnLine*, vol. 55, no. 10, pp. 1582–1587, 2022.

[32] F. Batool and C. Hennig, "Clustering with the average silhouette width," *Computational Statistics & Data Analysis*, vol. 158, p. 107190, 2021.

[33] A. Bhardwaj, "Silhouette coefficient validating clustering techniques," Towards Data Science, 2020, accessed Nov. 26, 2020. [Online]. Available: silhouette-coefficient-validating-clustering-techniques-e976bb81d10c

Dr. Rahmad Kurniawan received the BE(IT) degree in Informatics Engineering from the State Islamic University of Sultan Syarif Kasim Riau, Indonesia, in 2011, and the Master of Information Technology (MIT) in computer science from The National University of Malaysia, in 2014. From October 2011 – June 2022, he has been a Lecturer with the Department of Informatics Engineering, State Islamic University of Sultan Syarif Kasim Riau. In 2019 he received a Ph.D. from The National University of Malaysia. In 2022 as a Senior Lecturer in the Department of Computer Science, University of Riau. His current research interests include machine learning, expert system, data mining optimization, big data, and intelligent system.

Prof. Dr. Fitra Lestari is a Professor and Researcher in the Industrial Engineering Department at Universitas Sultan Syarif Kasim Riau, Indonesia. He finished his PhD project with a significant area in supply chain management at Universiti Teknologi Malaysia. He is currently an ASEAN Engineering Member and has published several articles in international journals about supply chain management, logistics and performance measurement. Prof. Fitra Lestari also leads the Emerging Technology and Continuous Improvement Lab at the Center of Islamic Data Science and Continuous Improvement (CIDSCI).

Assoc. Prof. Dr. Mawadi serves as a lecturer at the Faculty of Sharia and Law, Universitas Sultan Syarif Kasim Riau, Indonesia. Currently holding the position of Vice Dean II, Dr. Mawadi plays a pivotal role in assisting the Dean in overseeing the execution of activities related to general administration, finance, human resources, and facilities and infrastructure within the Faculty of Sharia and Law. In addition to these responsibilities, Dr. Mawadi also leads the Sharia Economics Lab at the Center of Islamic Data Science and Continuous Improvement (CIDSCI). His research and expertise are centered on Sharia economics, which implements the fundamental values and principles of Sharia derived from Islamic teachings. These Sharia principles, universal in their application are integral to various aspects of life, including economic and financial activities. Dr. Mawadi's work contributes significantly to the understanding and application of Sharia in the modern economic context.

Mrs. Tengku Nurainun is an Assistant Professor in Industrial Engineering Department at Universitas Sultan Syarif Kasim Riau, Indonesia. Currently, She is pursuing her PhD at Universiti Teknologi Malaysia with the main field of research is Quality Management. Her area of interest are Quality Management System, Integrated Management System, Halal Management System, and Small-Medium Enterprise (SME).

Prof. Dr. Abu Bakar Abdul Hamid is a seasoned researcher currently serving at Putra Business School in Malaysia. His research expertise spans both quantitative and qualitative social research methods, with a focus on various disciplines. Abu Bakar's skills and expertise are multifaceted, including Supply Chain Management (SCM), Logistics, and Management, with a particular emphasis on sectors such as the Automotive Industry, E-Business, and Manufacturing. He is adept in Performance Measurement, Operations Management, and Quality Management. His current role at Putra Business School involves engaging in cutting-edge research, contributing valuable insights and developments in his areas of expertise.

Tisha Melia is an Assistant Professor in the Department of Computer Science, University of Riau, Indonesia. She obtained her PhD degree from Boston University. Her research interest machine learning for biological data (Bioinformatics).