International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

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

Rahmad Kurniawan[1]*, Fitra Lestari[2], Mawardi[3], Tengku Nurainun[4], Abu Bakar Abdul Hamid[5]

[1] Department of Computer Science, Universitas Riau, Pekanbaru, Indonesia*
[2,4] 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
[5] Putra Business School, Universiti Putra Malaysia, Serdang, Malaysia

E-mail address: rahmadkurniawan@lecturer.unri.ac.id, fitra.lestari@uin-suska.ac.id, mawardi@uin-suska.ac.id, t.ainun@uin-suska.ac.id, abubakarhamid@iukl.edu.my

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: Cowhide plays a significant role in Indonesia's culinary, leather industries and caters to the preferences of a predominantly Muslim population that strongly emphasizes halal products. Regulatory authorities must comprehensively understand its characteristics to provide halal assurance to the diverse entities within Indonesia's leather industry effectively. This study employs unsupervised learning methods, specifically K-Means and Hierarchical clustering algorithms to analyze a dataset comprising 100 Cowhide The Small and Medium Enterprises (SMEs) Industries located in Garut Regency, West Java Province, Indonesia. This dataset encompasses 62 features that enable the clustering of cowhide industries based on halal risk factors. Experimental results indicate that the optimal number of clusters is m=4. The K-Means algorithm outperforms the Hierarchical clustering algorithm with a higher average silhouette score of 0.59 compared to 0.31 indicating its superior performance. Furthermore, the K-Means algorithm demonstrates exceptional stability in clustering the data, making it a robust choice for this analysis. The clustering outcomes of the Cowhide SMEs Industry provide valuable insights into the industry's characteristics, facilitating the efficient implementation of halal assurance measures. These findings hold substantial implications for the halal certification and assistance processes within the leather industry in Indonesia.

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. 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 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

E-mail:author's email

each SME 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.

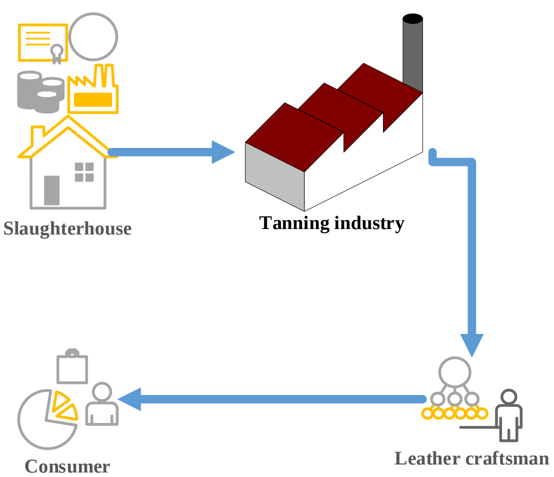


Figure 1. The previous investigation on the ideal supply chain within the leather industry

The assessment of halal compliance for leather craft products differs from that for cowhide-based food products. According to [3], 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 [4], which encompass the status of raw materials, processing methods, wholesomeness, and shared facilities for halal and non-halal products. Previous research has predominantly focused on identifying and managing halal risks in various contexts, such as the red meat industry [5] or the frozen food industry [6]. This study uses unsupervised learning approaches to cluster leather industries based on halal risk factors.

The clustering of leather SMEs industries serves as a means to gain insights into the characteristics of this sector. Industries that are clustered together are deemed to share similar characteristics. Traditional statistical risk assessment is considered inefficient when applied to extensive data rows and columns [7]. Consequently, competent authorities can efficiently provide halal assistance by identifying high-risk industries that do not adhere to halal principles in their production.

Machine learning has witnessed significant utilization across various domains for analyzing extensive datasets at risk [7]. 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 [8], [9], while deep learning methods have been employed to detect non-halal ingredients in food [10]. 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.

One widely used unsupervised learning method is clustering, which can group unlabeled data effectively, irrespective of the number of attributes [11]. Among the array of data analysis techniques, clustering is a method for categorizing data into groups sharing similar characteristics [12]. Clustering algorithms represent a potent approach for extracting valuable insights by grouping data points.

Both the K-means and Hierarchical Clustering algorithms have been employed to address problems across various domains, such as microarray efficiency [13], DNA sequence analysis [14], astronomy [15], and data visualization [16].

In a recent study, remarkable results have been achieved through clustering using the K-means algorithm to cluster 25 mammals [17]. 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 [18]. 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 [7], 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 [12].

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 [19]–[22].

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 [19]. The motivation for emphasizing the justification of clustering outcomes employing the Silhouette score is supported by several relevant studies. In a comparative study [19], 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 [19], Similarly, in a study on load profiles clustering methods [20], 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 [20].

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 [21] 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 [21].

Additionally, the Silhouette index has been explored in conjunction with the K-Harmonic Means method to group remote sensing datasets effectively [22]. This approach showcases the Silhouette index's ability to determine the correct number of clusters in scenarios with varying degrees of cluster overlap [22]. 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.

A study highlights the challenges of identifying compact and well-separated clusters within datasets, a task many clustering algorithms grapple with [23]. 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. Consequently, researchers have turned to cluster validity indices, such as Silhouette coefficients, to assess the number of well-separated and compact clusters [23].

Therefore, this study aims to propose a novel approach utilizing unsupervised learning algorithms to cluster the Leather SMEs Industry in Garut, Indonesia, based on halal risk factors. This approach provides valuable insights into the industry's characteristics and facilitates the efficient implementation of halal assistance, ultimately enhancing halal compliance within the cowhide industry. This study provides to the expanding knowledge of halal compliance within the cowhide industry and underscoring the significance of unsupervised learning algorithms in identifying and managing halal risks.

2. MATERIAL AND METHODS

Data for this study were collected from 100 SMEs located in Garut, West Java. These SMEs were selected as respondents to generate a dataset comprising 62 features and 100 records.

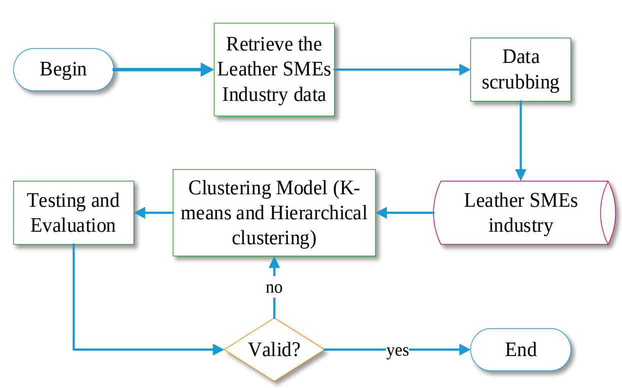


Figure 2. Methodology of an unsupervised learning algorithm for clustering the Cowhide SMEs Industry in Garut Regency, Indonesia

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

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

The Leather SMEs Industry in Garut can be clustered using the K-means and hierarchical clustering algorithms. Let δ_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:

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

The dataset $\beta = \{\beta_1, \ldots, \beta_n\}; n = 100$ comprises Cowhide SMES Industry names, types, number of employees, duration, halal certification, distribution, and 52 features from questionnaires. We tested various cluster sizes with $m = \{2,3,4\}$. To begin, we randomly select δ 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.

In the Hierarchical clustering algorithm we began the Hierarchical clustering algorithm by initializing the cluster $\beta = \{\beta_1, \ldots, \beta_n\}; n = 100$ be the set of Cowhide SMEs Industry in Garut data points containing of *Cowhide SMEs Industry names, types, Number of employees, durations, halal certification, distributions, and 62 features from questionnaires*, $\delta = \{\delta_1, \ldots, \delta_m\}$ 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):

$$\beta(\beta_i, \beta_n) = distance\ (\beta_i, \beta_n) = for\ all\ i, n\ such\ that\ i < n \qquad (2)$$

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.

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 [7]. The silhouette coefficient is calculated using the following formula 3 [24]:

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

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.

3. RESULTS AND DISCUSSION

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

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 *(a)*	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

Based on the results presented in Table 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.

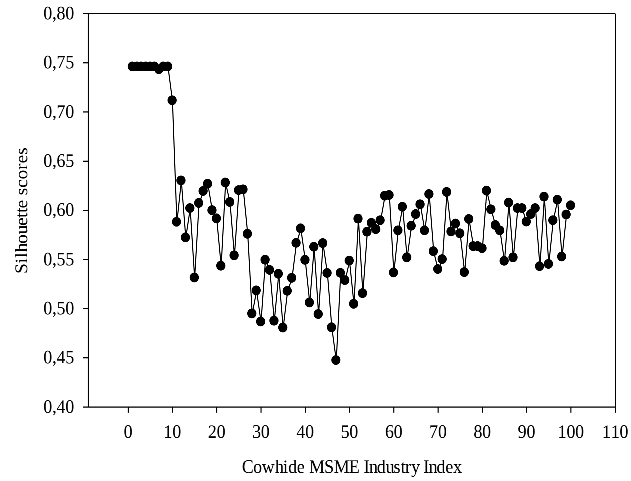


Figure 3. Silhouette scores when m=4 for the 100 Cowhide SMEs Industries in Garut using the K-Means clustering algorithm.

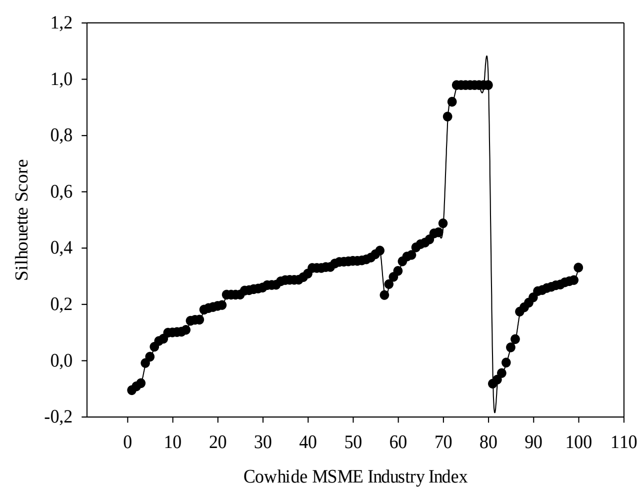
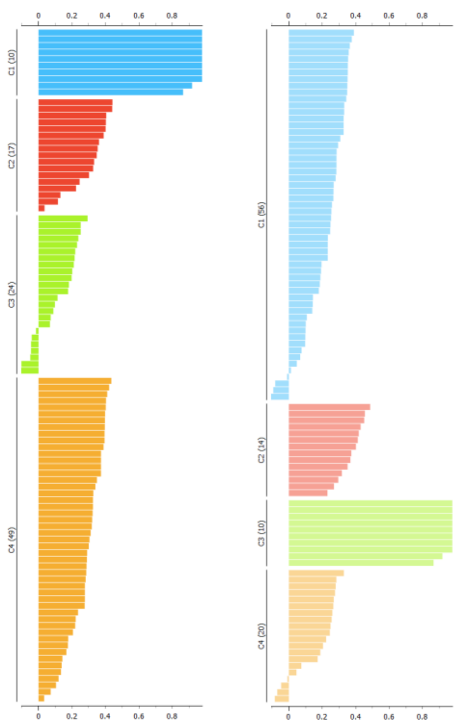


Figure 4. Silhouette scores when m=4 for the 100 Cowhide SMEs Industries in Garut using the Hierarchical clustering algorithm

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.



(a) *(b)*

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

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 analyze 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.

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

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

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

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.

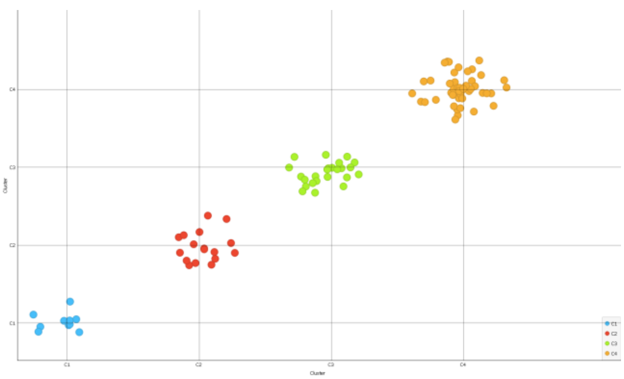


Figure 6. Visualization of clustering results using the K-means algorithm with the number of clusters set to $m = 4$

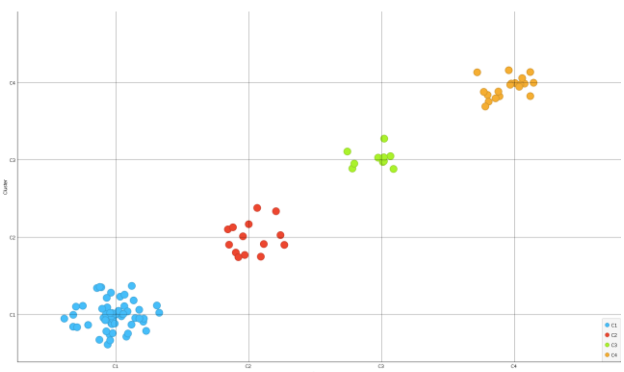


Figure 7. Visualization of clustering results using the algorithm with the number of clusters $m = 4$

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

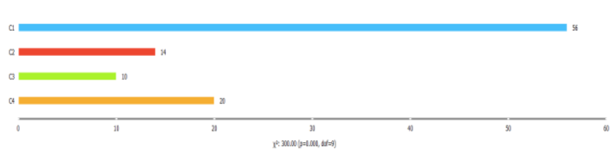


Figure 8. Visualization of box plot in the Hierarchical Clustering algorithm

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 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.

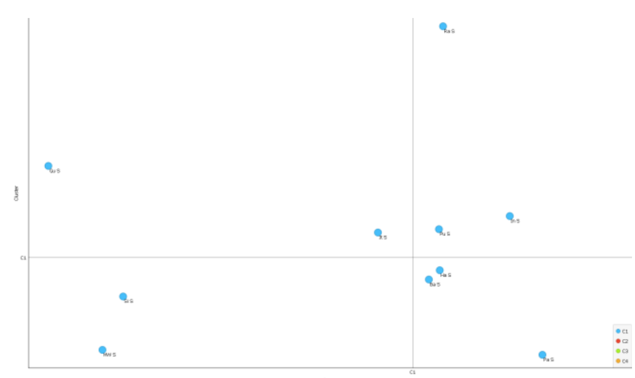


Figure 9. Assignment of multiple Cowhide SMEs Industries to the first cluster (C1) by the K-means algorithm when the number of clusters is set to $m = 4$

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.

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 three clusters (C1, C2, and C3). Identifying these clusters will simplify policymaking for SMEs and decision-makers, facilitating their efforts to support global halal standardization.

4. Conclusion

This study successfully applied K-means and Hierarchical clustering algorithms focusing on halal risk factors to the Cowhide SMEs Industry in Garut, West Java, Indonesia. The results effectively clustered industry data, offering valuable insights into halal compliance within the cowhide industry as a crucial aspect considering Indonesia's sizeable Muslim population and its focus on halal products.

Notably, the study identified the critical role of the tanning process in maintaining halal integrity, aligning with the Indonesian Ulama Council's Fatwa No. 56 of 2014 on permissible materials and procedures. The C4 cluster comprising industries involved in tanning was highlighted for its potential risk of non-halal substance use.

These findings provide a strategic framework for halal management and authorities to enhance halal compliance and risk assessment in similar industries. Deploying unsupervised learning algorithms facilitates a deeper understanding of industry-specific risks and supports the development of efficient halal certification processes.

Furthermore, the study underscores the importance of a halal supply chain in guiding policymakers and SMEs toward standardization and enhanced productivity. It suggests a systematic approach to identifying SMEs with varying halal risk levels, enabling targeted government support.

Expanding the dataset, exploring real-time monitoring systems, and engaging with industry stakeholders are recommended for future research. Such collaborations could develop proactive risk mitigation strategies, further strengthening the halal industry's growth and sustainability.

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,” *J. Islam. Mark.*, 2022, doi: 10.1108/JIMA-04-2021-0125.

[2] D. Amertaningtyas, “Pengolahan Kerupuk Rambak Kulit di Indonesia,” *J. Ilmu-Ilmu Peternak.*, vol. 21, no. 3, pp. 18–29, 2018, [Online]. Available: https://www.neliti.com/publications/101050/.

[3] I. Jaswir, E. A. Rahayu, N. D. Yuliana, and A. P. Roswiem, *Daftar Referensi Bahan-bahan yang Memiliki Titik Kritis Halal dan Substitusi Bahan Non-halal [The List of Reference Materials with Critical Halal Points and Substitutions of Non-halal Materials]*, vol. 4, no. 1. Komite Nasional Ekonomi dan Keuangan Syariah, 2020.

[4] 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 J. Sci. Res.*, 2022, doi: 10.1108/AGJSR-09-2022-0160.

[5] U. Maman, A. Mahbubi, and F. Jie, “Halal risk mitigation in the Australian–Indonesian red meat supply chain,” *J. Islam. Mark.*, vol. 9, no. 1, pp. 60–79, 2018, doi: 10.1108/JIMA-12-2015-0095.

[6] D. Kristanto and D. A. Kurniawati, “Development of halal supply chain risk management framework for frozen food industries,” *J. Islam. Mark.*, 2023, doi: 10.1108/JIMA-04-2022-0112.

[7] R. Kurniawan, S. N. H. S. Abdullah, F. Lestari, M. Z. A. Nazri, M. Akhmad, and N. Adnan, “Clustering and Correlation Methods for Predicting Coronavirus COVID-19 Risk Analysis in Pandemic Countries,” in *International Conference On Cyber & IT Service Management*, 2020, p. 138.

[8] 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. 83354–83362, 2019, doi: 10.1109/ACCESS.2019.2923275.

[9] R. Setik, R. M. T. R. L. Ahmad, and S. Marjudi, “Exploring Classification for Sentiment Analysis from Halal Based

Tweets," *2021 2nd Int. Conf. Artif. Intell. Data Sci. AiDAS 2021*, Sep. 2021, doi: 10.1109/AIDAS53897.2021.9574255.

[10] H. Fadhilah, E. C. Djamal, R. Ilyas, and A. Najmurrokhman, "Non-Halal Ingredients Detection of Food Packaging Image Using Convolutional Neural Networks," *Proceeding - 2018 Int. Symp. Adv. Intell. Informatics Revolutionize Intell. Informatics Spectr. Humanit. SAIN 2018*, pp. 131–136, Mar. 2019, doi: 10.1109/SAIN.2018.8673376.

[11] R. Baruri, A. Ghosh, R. Banerjee, A. Das, A. Mandal, and T. Halder, "An Empirical Evaluation of k-Means Clustering Technique and Comparison," *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019*, pp. 470–475, 2019, doi: 10.1109/COMITCon.2019.8862215.

[12] Nisha and P. J. Kaur, "Cluster quality based performance evaluation of hierarchical clustering method," *Proc. 2015 1st Int. Conf. Next Gener. Comput. Technol. NGCT 2015*, pp. 649–653, Jan. 2016, doi: 10.1109/NGCT.2015.7375201.

[13] T. S. Chen *et al.*, "A combined K-means and hierarchical clustering method for improving the clustering efficiency of microarray," *Proc. 2005 Int. Symp. Intell. Signal Process. Commun. Syst. ISPACS 2005*, vol. 2005, pp. 405–408, 2005, doi: 10.1109/ispacs.2005.1595432.

[14] L. Muflikhah, Widodo, W. F. Mahmudy, and Solimun, "DNA Sequence of Hepatitis B Virus Clustering Using Hierarchical k-Means Algorithm," *ICETAS 2019 - 2019 6th IEEE Int. Conf. Eng. Technol. Appl. Sci.*, Dec. 2019, doi: 10.1109/ICETAS48360.2019.9117565.

[15] H. Yu and X. Hou, "Hierarchical clustering in astronomy," *Astron. Comput.*, vol. 41, p. 100662, Oct. 2022, doi: 10.1016/j.ascom.2022.100662.

[16] L. M. C. Cabezas, R. Izbicki, and R. B. Stern, "Hierarchical clustering: visualization, feature importance and model selection," *Appl. Soft Comput.*, vol. 141, p. 110303, Jul. 2021, doi: 10.1016/j.asoc.2023.110303.

[17] Matt, "10 Tips for Choosing the Optimal Number of Clusters," *Towards Data Science*, 2019. https://towardsdatascience.com/10-tips-for-choosing-the-optimal-number-of-clusters-277e93d72d92 (accessed Apr. 16, 2020).

[18] F. Alqasemi, S. Al-Hagree, A. Aqlan, K. M. A. 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 2021*, 2021, pp. 1–4, doi: 10.1109/ICTSA52017.2021.9406544.

[19] T. Gupta and S. P. Panda, "Clustering Validation of CLARA and K-Means Using Silhouette DUNN Measures on Iris Dataset," *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019*, pp. 10–13, Feb. 2019, doi: 10.1109/COMITCON.2019.8862199.

[20] A. Bosisio *et al.*, "Performance assessment of load profiles clustering methods based on silhouette analysis," *21st IEEE Int. Conf. Environ. Electr. Eng. 2021 5th IEEE Ind. Commer. Power Syst. Eur. EEEIC / I CPS Eur. 2021 - Proc.*, 2021, doi: 10.1109/EEEIC/ICPSEUROPE51590.2021.9584629.

[21] G. Jin-Heng, L. Jia-Xiang, Z. Zhen-Chang, and L. Han-Yu, "CDBSCAN: Density clustering based on silhouette coefficient constraints," *Proc. - 2022 Int. Conf. Comput. Eng. Artif. Intell. ICCEAI 2022*, pp. 600–605, 2022, doi: 10.1109/ICCEAI55464.2022.00128.

[22] H. Mahi, N. Farhi, K. Labed, and D. Benhamed, "The Silhouette Index and the K-Harmonic Means algorithm for Multispectral Satellite Images Clustering.," *Proc. 2018 Int. Conf. Appl. Smart Syst. ICASS 2018*, Jul. 2018, doi: 10.1109/ICASS.2018.8652068.

[23] A. M. Bagirov, R. M. Aliguliyev, and N. Sultanova, "Finding compact and well-separated clusters: Clustering using silhouette coefficients," *Pattern Recognit.*, vol. 135, p. 109144, Mar. 2023, doi: 10.1016/J.PATCOG.2022.109144.

[24] A. Bhardwaj, "Silhouette Coefficient Validating clustering techniques," *Towards Data Sience*, 2020. https://towardsdatascience.com/silhouette-coefficient-validating-clustering-techniques-e976bb81d10c (accessed Nov. 26, 2020).

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 10 h Dat 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.