THE USE OF WEIGHTED K-MEANS CLUSTERING AND ANALYTICAL HIERARCHY PROCESS IN HALAL CERTIFICATION READINESS FOR MSMES IN YOGYAKARTA CITY

UNDERGRADUATE THESIS

Submitted to the International Undergraduate Program in Industrial Engineering in Partial Fulfilment of Requirement for the Degree of Sarjana Teknik at the Faculty of Industrial Technology Universitas Islam Indonesia

Name : Naoval Harief

Student Number : 20522284

UNDERGRADUATE PROGRAM IN INDUSTRIAL ENGINEERING
FACULTY OF INDUSTRIAL TECHNOLOGY
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2025

AUTHENTICITY STATEMENT

For the sake of Allah *Subhanahu Wata'alaa*, I admit this work is the result of my own work except for the excerpts and summaries from which I have explained the source. If in the future, it turns out that my confession is proven to be untrue and violates the legal regulations in the paper and intellectual property rights, then I am willing to get a diploma that I have received to be withdrawn by the Universitas Islam Indonesia.

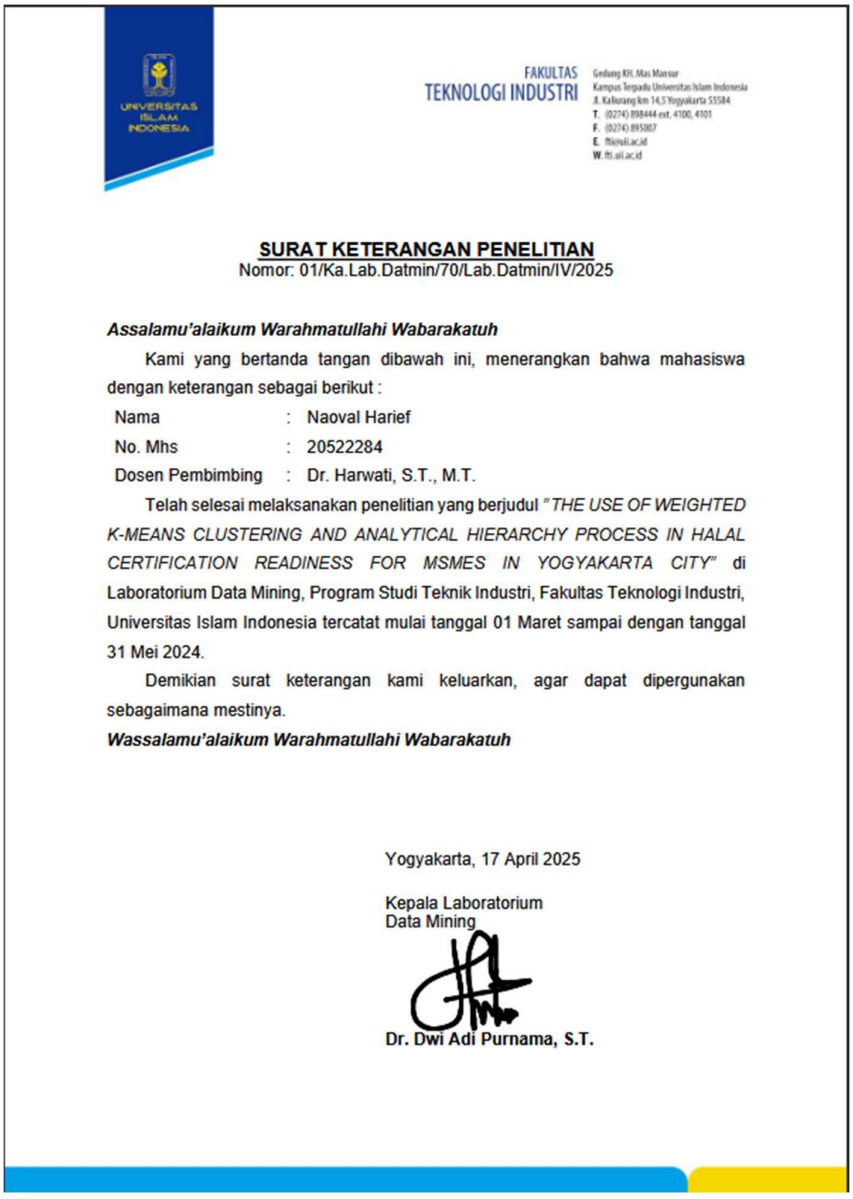
Yogyakarta, April 13, 2025



Naoval Harief

20522284

RESEARCH COMPLETION LETTER



UNIVERSITAS ISLAM INDONESIA

FAKULTAS TEKNOLOGI INDUSTRI

Gedung KH. Mas Mansur
Kampus Terpadu Universitas Islam Indonesia
Jl. Kaliurang km 14,5 Yogyakarta 55584
T. (0274) 898444 ext. 4100, 4101
F. (0274) 895007
E. ftiuii.ac.id
W. fti.uii.ac.id

SURAT KETERANGAN PENELITIAN
Nomor: 01/Ka.Lab.Datmin/70/Lab.Datmin/IV/2025

Assalamu'alaikum Warahmatullahi Wabarakatuh

Kami yang bertanda tangan dibawah ini, menerangkan bahwa mahasiswa dengan keterangan sebagai berikut :

Nama	:	Naoval Harief
No. Mhs	:	20522284
Dosen Pembimbing	:	Dr. Harwati, S.T., M.T.

Telah selesai melaksanakan penelitian yang berjudul "*THE USE OF WEIGHTED K-MEANS CLUSTERING AND ANALYTICAL HIERARCHY PROCESS IN HALAL CERTIFICATION READINESS FOR MSMES IN YOGYAKARTA CITY*" di Laboratorium Data Mining, Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia tercatat mulai tanggal 01 Maret sampai dengan tanggal 31 Mei 2024.

Demikian surat keterangan kami keluarkan, agar dapat dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Yogyakarta, 17 April 2025

Kepala Laboratorium
Data Mining

Dr. Dwi Adi Purnama, S.T.

SUPERVISOR APPROVAL SHEET

THE USE OF WEIGHTED K-MEANS CLUSTERING AND ANALYTICAL HIERARCHY PROCESS IN HALAL CERTIFICATION READINESS FOR MSMES IN YOGYAKARTA CITY

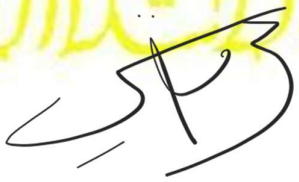
UNDERGRADUATE THESIS

Written by:

Name : Naoval Harief
Student Number : 20522284

Yogyakarta, April 13, 2025

Supervisor,



(Dr. Harwati, S.T., M.T.)

EXAMINERS' APPROVAL PAGE

THE USE OF WEIGHTED K-MEANS CLUSTERING AND ANALYTICAL HIERARCHY PROCESS IN HALAL CERTIFICATION READINESS FOR MSMES IN YOGYAKARTA CITY

UNDERGRADUATE THESIS

Written by:

Name : Naoval Harief

Student Number : 20522284

Has been defended before the board of examiners in partial fulfillment of the requirement for Bachelor Degree in Industrial Engineering Department at the Faculty of Industrial Technology Universitas Islam Indonesia

Yogyakarta, April 13, 2025

Board of Examiners

Dr. Harwati, S.T., M.T.	
Chairman	
Dr. Drs. Imam Djati Widodo, M.Eng.Sc.	
Member I	
Dr. Agus Mansur, S.T., M.Eng. Sc.	
Member II	

Acknowledged by,
Head of Undergraduate Program in Industrial Engineering
Faculty of Industrial Technology
Universitas Islam Indonesia

Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM
NIK. 015220101

DEDICATION PAGE

This Undergraduate Thesis, which took a lot of time and hard work. I dedicate this thesis to my beloved parents, Father Tarip Suharwito, my Mother Sri Hastuti, and My Sister Ifa, whose unwavering love, support, and prayers have been my guiding light throughout this journey. Also extend my gratitude to all my friends, whose camaraderie, encouragement, and motivation have made this achievement possible.

May this work be a testament to the love and appreciation I have for all of you.

MOTTO

"For indeed, with hardship comes ease, indeed, with hardship comes ease."
(Q.S AL-Insyirah: 5-6)

"*Make it happen, don't wait"*

PREFACE

Bismillahirrahmanirrahim
Assalamu'alaikum Warahmatullahi Wabarakatuh

Alhamdulillahirabbil'alamin, all praise be to Allah SWT, the Most Generous, whose mercy and grace have enabled the completion of this undergraduate thesis titled *"The Use of Weighted K-Means Clustering and Analytical Hierarchy Process in Halal Certification Readiness for MSMEs in Yogyakarta City."* Peace and blessings be upon the Prophet Muhammad SAW, who guided humanity to the light of knowledge. This work would not have been possible without the support of many. With sincere gratitude, the author extends appreciation to all who have contributed morally and materially:

1. Prof. Dr. Ir. Hari Purnomo., M.T., IPU, ASEAN.Eng as the Dean of the Faculty of Industrial Technology, Universitas Islam Indonesia.
2. Dr. Drs. Imam Djati Widodo, M.Eng.Sc., as the Head of the Department of Industrial Engineering, Faculty of Industrial Technology, Universitas Islam Indonesia.
3. Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., as the Head of the Industrial Engineering Study Program, Faculty of Industrial Technology, Universitas Islam Indonesia.
4. Dr. Ir. Agus Mansur, S.T., M.Eng.Sc., IPU**.**, as the Vice Dean for Resource Management, Faculty of Industrial Technology, Universitas Islam Indonesia.
5. Dr. Harwati, S.T., M.T., as the Undergraduate Thesis Supervisor, for her patience and sincere guidance throughout the research process.
6. Ms. Devy Nurrahmah, S.Kom., as the Academic Staff of the Industrial Engineering International Program, Faculty of Industrial Technology, who has always been helpful in providing academic information promptly.
7. My beloved family: Mr. Tarip Suharwito, S.T., M.M., and Mrs. Sri Hastuti, for their endless love, support, and prayers. To my sister, Dr. Rifdha Haifa Salsabila, and All my friends from the 2020 Industrial Engineering Study Program, whose companionship and support meant so much—even though I cannot mention each of you by name. thank you for your encouragement and motivation during difficult times.

This thesis was prepared as part of the requirements for the completion of the Bachelor's Degree in Industrial Engineering at the Faculty of Industrial Technology, Universitas Islam Indonesia.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Yogyakarta, April 13, 2025

Nuval Harief

NIM.20522284

ABSTRACT

This study investigates the readiness of Micro, Small, and Medium Enterprises (MSMEs) in Yogyakarta City, Indonesia, for halal certification using a novel approach combining Weighted K-Means Clustering and Analytical Hierarchy Process (AHP). As Indonesia aims to become a global halal hub, understanding the preparedness of MSMEs for halal certification is crucial. The research employs AHP to determine the relative importance of factors influencing halal certification readiness, which are then used as weights in a K-Means clustering algorithm. Data collected from MSMEs in Yogyakarta City, encompassing financial resources, current production processes, halal knowledge, and market demand, are analyzed using this weighted clustering approach. The resulting clusters reveal distinct groups of MSMEs with varying levels of readiness for halal certification. This segmentation provides valuable insights for policymakers and support organizations to develop targeted interventions and strategies. The study contributes to the growing body of literature on halal certification and offers a methodological framework for assessing MSME readiness in other contexts. The findings have significant implications for facilitating the transition to mandatory halal certification while supporting the growth and competitiveness of local businesses in the global halal market.

Keywords: Halal Certification, MSMEs, Weighted K-Means Clustering, Analytical Hierarchy Process, Indonesia

TABLE OF CONTENT

AUTHENTICITY STATEMENT ii
RESEARCH COMPLETION LETTER iii
SUPERVISOR APPROVAL SHEET iv
EXAMINERS' APPROVAL PAGE v
DEDICATION PAGE vi
MOTTO vii
PREFACE viii
ABSTRACT ix
TABLE OF CONTENT x
LIST OF TABLES xii
LIST OF FIGURES xiii
CHAPTER I INTRODUCTION 1
1.1 Background 1
1.2 Problem Formulation 2
1.3 Research Objective 3
1.4 Significance of Research 3
1.5 Research Limitation 3
1.6 Systematic Research 4
CHAPTER II LITERATURE REVIEW 6
2.1 Literature Review 6
2.1.1 Micro Small and Medium Enterprises 6
2.1.2 Knowledge Discovery in Database (KDD) 6
2.1.3 Data Mining 7
2.1.4 Multi Criteria Decision Making (MCDM) 7
2.1.5 Analytical Hierarchy Process (AHP) 7
2.1.6 Clustering 8
2.1.7 Weighted Clustering K-Means 8
2.1.8 Elbow 8
2.1.9 Gap Statistic 9
2.1.10 Davies Bouldin Index 9
2.1.11 Dunn Index 9
2.1.12 Calinski Harabasz Index 10
2.1.13 Silhouette 11
2.1.14 Python 11
2.2 Theoretical Basis 11
CHAPTER III RESEARCH METHODOLOGY 29
3.1 Research Objective 29
3.2 Research Subject 29
3.3 Data Sources 29
3.3.1 Primary Data 30
3.3.2 Secondary Data 30
3.4 Data Collection Method 30
3.4.1 Literature Study 30
3.4.2 Observation 31

3.4.3 Questionnaire 32
3.5 Data Analyzing Method 33
3.6 Research Flowchart 35
CHAPTER IV DATA COLLECTION AND PROCESSING OF SYSTEM DEVELOPMENT 37
4.1 Introduction 37
4.2 Data Collection 37
4.3 Pre-Processing Data 38
4.4 Data Transformation 38
4.4.1 Transformation Rules 38
4.4.2 Analytical Hierarchy Process Data Transformation 40
4.4.3 Clustering Data Transformation 43
4.5 Analytical Hierarchy Process Data Processing 44
4.5.1 Arithmetic Means Theory 44
4.5.2 Consistency Test 45
4.5.3 Weighting Result 47
4.6 Clustering Data Processing 48
4.6.1 Data Standardization 48
4.6.2 Determining the Cluster 49
4.6.3 Optimal Cluster Analysis with a Hybrid Approach 53
4.6.4 Cluster Visualization 57
4.6.5 K-Centroid Value and Final Cluster 59
4.6.6 Cluster Profile Group 61
4.7 Comparison Between the Weighted and Unweighted K-Means Clustering 63
4.7.1 Dataset 63
4.7.2 T – test Calculation Result 64
CHAPTER V RESULT AND DISCUSSION 67
5.1 The Comparison Between the Normal K-Means and the Weighted K-Means 67
5.2 K-Means Weighted Result 68
5.3 Cluster Profiling 69
5.4 Development Strategy Based on Halal Certification for MSMEs 73
CHAPTER VI CONCLUSION AND SUGGESTION 76
6.1 Conclusion 76
6.2 Suggestion 77
REFERENCES 79
APPENDICES A-1
A. Questionnaire (Google Form) A-1
B. Clustering MSMEs Data Recapitulation B-3
C. AHP (Pairwise Comparison) Data C-5
D. Gap statistic Python Coding Script D-8

LIST OF TABLES

Table 2.1 Theoretical Basis Recent Researcher 12
Table 3.2 Research Literature Study 30
Table 3.3 Questionnaire Question Draft 32
Table 4.4 Variable Attribute Code 38
Table 4.5 Transformation Rules 39
Table 4.6 Expert 1 Pairwise Comparison Matrix 40
Table 4.7 Expert 2 Pairwise Comparison Matrix 41
Table 4.8 Expert 3 Pairwise Comparison Matrix 41
Table 4.9 Expert 4 Pairwise Comparison Matrix 42
Table 4.10 Expert 5 Pairwise Comparison Matrix 43
Table 4.11 Transformed Data 43
Table 4. 12 Criterias Description 44
Table 4.13 Arithmetic Means Theory Calculation Result 45
Table 4.14 Data Normalized 46
Table 4.15 Consistency Test 46
Table 4.16 Ratio Index Table 47
Table 4.17 Criteria Weight Value 48
Table 4.18 Standardize K-Means Clustering 48
Table 19 Standardize Weighted Clustering K-Means 49
Table 20 Cluster Member of Unweighted K-Means Clustering 56
Table 21 Cluster Member of Weighted K-Means Clustering 57
Table 22 Centroid Value of Unweighted K-means 59
Table 23 Centroid Value of Weighted K-Means 59
Table 24 Gap Statistic Unweighted 60
Table 25 Gap Statistic Weighted 60
Table 26 Rounding Value of Unweighted K Means 61
Table 27 Transformed Rounding Value of Unweighted K Means 61
Table 28 Rounding Value of Weighted K Means 62
Table 29 Transformed Rounding Value of Weighted K Means 62
Table 30 Unweighted K-Means Clustering Result 63
Table 31 Weighted K-Means Clustering 64
Table 32 T-test Calculation 65
Table 33 Weight on Attributes 69
Table 34 Weighted Cluster Profiling Result 70
Table 35 Cluster Members 70

LIST OF FIGURES

Figure 3.1 Research Flowchart 35
Figure 4.2 Gap Statistic Unweighted 50
Figure 4.3 Gap Statistic Weighted 50
Figure 4.4 Unweighted Calinski Harabasz 51
Figure 4.5 Weighted Calinski Harabasz 52
Figure 4.6 Python Using Google Colab 55
Figure 4.7 UMAP Visualization for Unweighted K-means 58
Figure 4.8 UMAP Visualization for Weighted K-means 58

CHAPTER I
INTRODUCTION

1.1 Background

The global halal market is rapidly expanding, spurred by a projected Muslim population growth of 1.5% annually, nearly double the rate of non-Muslims. By 2030, Muslims are expected to represent 26.4% of the global population. This demographic shift fuels increased demand for halal certification, enhancing international trade opportunities. Indonesia is a key player in this landscape, having climbed from 4th to 3rd place in the State of the Global Islamic Economy (SGIE) rankings from 2022 to 2023. The country excels in sectors like Islamic banking and halal industries, holding first place in the Global Muslim Travel Index (GMTI) 2023 and third in the Global Islamic Finance Report (GIFR) 2022. However, some sectors, such as halal pharmaceuticals and media, have seen declines. Indonesia's recognition in 2023 with the Impactful Achievement in Islamic Economics award underscores its growing influence.

The enactment of Law Number 33 of 2014 mandates halal certification for all products by October 17, 2024. This regulation aims to bolster consumer trust but raises concerns among Micro, Small, and Medium Enterprises (MSMEs) about compliance costs impacting their competitiveness. MSMEs are vital to Indonesia's economy, comprising over 64.2 million businesses and employing around 97.2 million people. They represent 61.4% of all business entities and have shown resilience during economic downturns. The halal sector alone is projected to contribute 25.39% to GDP, with a growth rate of 4.37% in Q3 2023. Despite these opportunities, challenges persist. Approximately 50% of the population lacks awareness of halal principles, and there is a shortage of trained professionals in the halal industry. Many halal products struggle to differentiate themselves from conventional options, limiting market appeal. Additionally, coordination inefficiencies among stakeholders and political risks further complicate growth. The government has introduced initiatives, such as the Halal Product

Guarantee Agency (BPJPH) and financial assistance for certification, yet the adoption rate among MSMEs remains low, highlighting the need for targeted support. In Yogyakarta City a region well-known for its vibrant local economy and strong MSME sector the Department of Industry, Cooperatives, and MSMEs (Perinkop UKM) has been actively encouraging food and beverage businesses to obtain halal certification ahead of the 2024 deadline. However, many local MSMEs are still unprepared due to limited awareness, lack of knowledge, and perceived financial burdens. To address this, the city government is intensifying educational campaigns and providing halal certification assistance through both self-declared schemes and regular certification programs. Despite these efforts, the dynamic and informal nature of MSMEs in Yogyakarta makes it difficult to monitor readiness levels and tailor support effectively.

This research aims to assess the readiness of MSMEs in Yogyakarta for halal certification using a hybrid approach combining Weighted K-Means Clustering and the Analytical Hierarchy Process (AHP). AHP is used to assign weights to key factors such as knowledge, awareness, information access, cost, perception, and technology readiness based on expert input. These weighted variables are then used in the clustering process to group MSMEs by their level of readiness. By identifying patterns and readiness levels, this study seeks to provide actionable insights for policymakers and stakeholders to enhance MSMEs' integration into the halal market and improve their competitiveness both locally and globally.

1.2 Problem Formulation

Below is the problem based on the background above:

1. What are the results of grouping Yogyakarta MSMEs based on existing criteria variables using the weighted clustering method?
2. What is the best form of suggestion for the group formed to assist the community and government that contribute to the mandatory implementation of halal certification for MSMEs?

1.3 Research Objective

Below are the research objectives based on the background above:

1. Identifying the weighted clustering analysis with the K-Means algorithm in determining the influence factors on MSMEs in Yogyakarta City that affect the obtaining of halal certification.
2. Determine the results of the analysis of weighted clustering datamining with the K-Means algorithm in determining the grouping of MSMEs based on the level of readiness which refers to the existing criteria variables.

1.4 Significance of Research

This research aims to address the critical gap in understanding the diverse readiness levels of MSMEs for halal certification. By identifying key factors, developing a weighted clustering method, and proposing targeted strategies, this study will provide valuable insights for policymakers, stakeholders, and support organizations:

1. Enhance the effectiveness of halal certification implementation for MSMEs.

2. Reduce the burden on MSMEs and streamline the certification process.

3. Accelerate the adoption of halal certification among MSMEs, boosting their competitiveness in the global halal market.

1.5 Research Limitation

Below are the limitations of the research based on the halal certification adoption among MSMEs:

1. Limited sample size: Due to resource constraints, researcher may only be able to survey a small number of MSMEs, which may not be representative of the entire population.
2. Self-reported data: The data used in the research may be self-reported by the MSMEs, which may be subject to bias and inaccuracies.

3. Lack of longitudinal data: Most research on halal certification adoption among MSMEs is cross-sectional, which means that it only captures data at a single point in time. This makes it difficult to determine the long-term impact of halal certification on MSMEs.
4. Limited scope: The research may only focus on certain aspects of halal certification adoption, such as the cost or time required to obtain certification, and may not consider other important factors, such as consumer demand or market access.
5. Limited generalizability: The findings of the research may only apply to the specific context of Yogyakarta City and may not be generalizable to other regions or countries.
6. Data availability: There may be limited data available on halal certification adoption among MSMEs, particularly in developing countries, which may make it difficult to conduct comprehensive research.
7. Lack of standardization: There is no universally accepted standard for halal certification, which may make it difficult to compare the results of different studies and to draw conclusions about the effectiveness of halal certification.
8. Cultural and religious differences: The interpretation and implementation of halal certification may vary across different cultures and religions, which may affect the adoption of halal certification among MSMEs.

1.6 Systematic Research

The research is structured into several chapters, and each chapter will be described as follows:

CHAPTER 1 INTRODUCTION

An introduction is the opening section of a research paper, article, or essay that sets the stage for the entire document. It serves as a gateway to the topic, providing readers with a clear understanding of the subject matter, its significance, and the purpose of the research.

CHAPTER II LITERATURE REVIEW

A literature review is a critical analysis of the existing research on a particular topic or issue. It is a comprehensive and systematic examination of the existing body of knowledge on a specific research question or topic, including theoretical and methodological contributions.

CHAPTER III RESEARCH METHODOLOGY

Research methodology refers to the systematic and structured approach used to collect, analyze, and interpret data in a research study. It involves the selection of research design, sampling strategy, data collection methods, data analysis techniques, and other procedures to ensure the validity, reliability, and generalizability of the research findings.

CHAPTER IV DATA COLLECTING AND PROCESSING

Data collecting and processing are crucial steps in the research methodology that involve gathering and transforming raw data into a usable format for analysis.

CHAPTER V DISCUSSION

The discussion section is a critical component of a research paper, where you interpret the results, relate them to the broader context, and draw conclusions based on the findings.

CHAPTER VI CONCLUSIONS AND SUGGESTIONS

A conclusion is a crucial part of a research paper, summarizing the main findings and reiterating the significance of the study.

REFERENCES

APPENDIX

CHAPTER II
LITERATURE REVIEW

2.1 Literature Review

2.1.1 Micro Small and Medium Enterprises

Micro, Small, and Medium Enterprises (MSMEs) are defined as productive economic businesses owned by individuals or business entities, categorized based on their scale and financial criteria. In Indonesia, MSMEs are classified into four main types: Micro Enterprises, which are small-scale operations owned by individuals; Small Enterprises, which operate independently and are not branches of larger companies; Medium Enterprises, which are standalone businesses with specific asset or sales thresholds; and Large Enterprises, characterized by significant net worth or sales exceeding that of medium enterprises. This classification not only facilitates a better understanding of the business landscape but also aids in the development and support of these enterprises, which play a vital role in the economy (Arifin et al., 2021). MSMEs (Micro, Small, and Medium Enterprises) are businesses that are owned and operated by individuals, households, or entities that fall within certain size limitations. These businesses are typically smaller in scale compared to large corporations. They play a crucial role in various economies, providing employment opportunities, promoting economic growth, and contributing to innovation and entrepreneurship. MSMEs often operate in diverse sectors, from retail and services to manufacturing and agriculture.

2.1.2 Knowledge Discovery in Database (KDD)

Knowledge discovery in databases is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. Knowledge Discovery in Databases (KDD) is an automatic, exploratory analysis and modeling of large data repositories. KDD is the organized process of identifying valid, novel, useful, and understandable patterns from large and complex data sets. Data Mining (DM) is the core of the KDD process,

involving the inferring of algorithms that explore the data, develop the model and discover previously unknown patterns (Markusheski et al., 2019.)

2.1.3 Data Mining

Data mining is the process of sorting through large data sets to identify patterns and relationships that can help solve business problems through data analysis. Data mining techniques and tools help enterprises to predict future trends and make more informed business decisions. Datamining is the term used to dig up hidden knowledge in a database. Data mining is a semi-automatic process using statistical techniques , mathematics, artificial intelligence, and machine learning to extract and identify the information of potential and useful knowledge saved in a big database (Hartama et al., 2019). An algorithmic approach to finding new, relevant, and fascinating information in databases is known as data mining (DM). Mathematical statistics, probability distributions, statistical inference, as well as neural networks are all applied disciplines of study in DM algorithms. It is possible to utilize some of these areas' methodologies to uncover hidden relationships among data, which may then be used to build models that anticipate behavior or define certain common characteristics of the things being examined (Abdulkadium et al., 2022).

2.1.4 Multi Criteria Decision Making (MCDM)

Multi-criteria decision making is an approach or a method for making a decision based on the alternatives or options from the multi-criteria. Multi criterion Decision-Making (MCDM) tools are generally applied in arriving at an optimum decision when faced with multiple alternatives having multiple conflicting and non-commensurable decision criteria. The technique is a well-known tool for solving complex real-life problems due to its intrinsic ability to judge diverse alternatives with reference to various decision criteria in order to choose to best alternative (Emovon & Oghenenyerovwho, 2020).

2.1.5 Analytical Hierarchy Process (AHP)

The analytical hierarchy process is a method for managing and analyzing complex decisions, using mathematics and psychology.

The Analytical Hierarchy Process (AHP) has been a favorite tool of research experts from various fields such as engineering, technology, manufacturing, production, social sciences, etc. It has proved to be a reliable and efficient technique (Khan & Ali, 2020).

The AHP (Analytic Hierarchy Process) method was developed in the 1980s by Saaty. It is a systematic decision-making method that includes both qualitative and quantitative techniques. It is useful for obtaining a single assessment value based on different indicators or criteria. It simplifies the process of decision making by subdividing a complex problem into a series of structured steps where each element in the hierarchy of criteria is supposed to be independent of all the others (Benmoussa et al., 2019).

2.1.6 Clustering

Clustering is a method of grouping data. Clustering is a process of grouping data into several groups of clusters so that the data in each cluster has a certain type of similarity, which is maximum, and the data between clusters has a minimum similarity. Clustering is a process of partitioning a set of data objects into subsets called clusters. Clustering is a method of unsupervised machine learning used to explore and analyze datasets without relying on predefined labels or response variables. In this case, clustering is employed to analyze factors related to student performance in university exams. The primary advantage of clustering is its ability to group similar observations based on their characteristics. Given that the dataset includes various factors influencing student performance, clustering helps identify patterns or groupings within the data. These groupings can reveal relationships between different factors and their interactions affecting student performance (Aggarwal & Sharma, 2019).

2.1.7 Weighted Clustering K-Means

Weighted clustering K-Means works the same as standard K-Means clustering. The only difference would be that instead of just calculating centroid points based on the mean distances. The weighted average should be used. Thus, the bigger the weight of the data point, the nearer the centroid will be pulled. In the AHP method, each criterion has a weight in this case, the weight value is obtained from the results of the interview with the Department of Cooperatives and MSME in the MSME sector (Sukmadewanti et al., 2018).

2.1.8 Elbow

The elbow method is a graphical method for finding the optimal K value in a K-means clustering algorithm. The elbow graph shows the within-cluster-sum-of-square values on the y-

axis corresponding to the different values of K. The elbow method is suitable for relatively small k values.

The elbow method calculates the squared difference of different k values. As the k value increases, the average distortion degree becomes smaller. The number of samples contained in each category decreases, and the samples are closer to the center of gravity. As the k value increases, the position where the improvement effect of the distortion degree decreases the most is the k value corresponding to the elbow (Cui, 2020).

2.1.9 Gap Statistic

The Gap statistic (Sinaga & Yang, 2020) is a cluster validity measure based on a statistical hypothesis test. The gap statistic compares the change in within-cluster dispersion with that expected under an appropriate reference null distribution at each value c. The optimal number of clusters is the smallest c.

2.1.10 Davies Bouldin Index

Davies Bouldin Index is used to measure the quality of the clustering results (Kristanto et al., 2023) Davies-Bouldin Index (DBI) is a metric used to evaluate the quality of clustering results. It is the ratio between the intra-cluster distance and the inter-cluster distance. The DBI value indicates how well the clusters are formed, with lower values indicating better clustering results.

$$DB = \frac{1}{n}\sum_{i=1}^{n} max_{j \neq i}\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right)$$

where n is the number of clusters and σ i is the average distance of all points in cluster i from the cluster centroid c i. The DB index captures the intuition that clusters that are: (1) well-spaced from each other and (2) themselves very dense are likely a ‘good’ clustering (Abdulhafedh, 2021).

2.1.11 Dunn Index

The Dunn Index captures the same idea as the DB Index: it gets better when clusters are well-spaced and dense. But the Dunn Index increases as performance improves.

In order to compare the results obtained from both the Hierarchical clustering and K-means (Abdulhafedh, 2021).

$$D = \frac{min_{1\le i<j\le n} d(i,j)}{max_{1\le k\le} d'(k)}$$

where i, j and k are indices for clusters, d measures the inter-cluster distance and d' measures the intra-cluster difference.

2.1.12 Calinski Harabasz Index

The Calinski-Harabasz Index is an evaluation index based on the degree of dispersion between clusters and clusters (Wang & Xu, 2019). Defined as follows:

$$CH(K) = \frac{B(K)(N-K)}{W(K)(K-1)}$$

$$B(K) = \left(\sum_{k=1}^{K} a_k \|\overline{x_k} - \bar{x}\|^2\right)$$

$$W(K) = \left(\sum_{k=1}^{K} \sum_{c(j)=k} \left\|\overline{x_j} - \overline{x_k}\right\|^2\right)$$

Among them: K is the corresponding number of clusters, B(K) is the inter-cluster divergence, also called the inter-cluster covariance, W(K) is the intra-cluster divergence, also called the intra-cluster covariance, and N is the number of samples. The larger the B(K) is, the higher the degree of dispersion between clusters is. The smaller the W(K) is, the closer the relationship is in the cluster. The higher the ratio, the larger the value of the CH index is, that is, the better the clustering effect is.

2.1.13 Silhouette

The Silhouette Index obtains the optimal clustering number by the difference between the average distance within the cluster and the minimum distance between the clusters (Wang & Xu, 2019).

That is, the optimal clustering effect, which is defined as follows:

$$\bar{S} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{b(i) - a(i)}{\max\{a(i), b(i)\}}\right)$$

among them: a(i) represents the average distance of sample i to other samples in the cluster, b(i) represents the minimum distance of the sample from sample i to the other clusters.

2.1.14 Python

Python is a high-level, general-purpose programming language that emphasizes code readability with the use of significant indentation. It is an interpreted, object-oriented language with dynamic semantics, making it a versatile language in various fields. Python is often used for web development, scientific computing, data analysis, artificial intelligence, and more. Python is a language of open-source programming language that is both good and easy to read. It was made in 1991 by a programmer named Guido van Rossum. Python is named after the Flying Circus show by Monty Python. The show's jokes include many examples and tutorials. Python is an interpret language. For running, there is no need to compile interpreted languages. On any computer that it can run on itself, a program called an interpreter runs Python code. This means that if the programmer has to change the code, they can quickly see the results. This also means that Python is slower than a compiled language such as C because it does not run machine code directly (Palariya, 2019).

2.2 Theoretical Basis

Theoretical basis is a research method that entails formulating ideas or generalizations by analyzing specific facts or patterns.

Table 2.1 Theoretical Basis Recent Researcher

Researcher/Year	Title	Method	Output
Ari Novriansyah Dwi Ely Kurniawan	Development of Fish Marketing Application for UMKM and Batam Fishermen with User Centered Design (UCD) Approach and Usability Testing	Laravel framework for application development. User Centered Design (UCD) approach to involve users in application design. COD (Cash on Delivery) payment model to overcome people's hesitation in online transactions. Evaluation of usability level using SUS (System Usability Scale) scale.	Designed fish marketing application. SUS scores of 57, 58, indicate that the application is functionally sound and acceptable to users.
Rizka Hafsari Gunadi Widi Nurcahyo	Application of Fuzzy C-Means on Technology Adoption of Micro, Small and Medium Enterprises.	Fuzzy C-Means (FCM) to determine the level of technology adoption of MSMEs. Criteria used: - Name of MSME	The results of data analysis in iteration 1 show that the Fuzzy C-Means method has not produced the expected error value, and the objective function

Researcher/Year	Title	Method	Output
		- Year of business start - Type of business - Number of workers - Capital - Assets - Turnover	obtained is 439.8768.
Dwi Utari Iswavigra Lova Endriani Zen Okfalisa Hafizah Hanim	Marketing Strategy for MSMEs with CRISP-DM Clustering and Promotion Mix Using the K-Medoids Method	CRISP-DM Clustering K-Medoids	This research resulted in 3 clusters of MSMEs with different characteristics, which allowed researchers to formulate targeted and effective promotional strategies for each cluster. 3 MSME Clusters: - Cluster 1: 25 MSMEs, dominance of fashion businesses, average turnover and assets

Researcher/Year	Title	Method	Output
			of Rp.1,000,000 - Rp.5,000,000. - Cluster 2: 39 MSMEs, dominated by culinary businesses, average turnover and assets of Rp.1,000,000 - Rp.5,000,000. - Cluster 3: 7 MSMEs, dominated by businesses outside the other 6 business types, average assets ≥ Rp.30,000,000, turnover Rp.21,000,000 - Rp.25,000,000.
Tiara Jelita Relita Buaton Magdalena Simajuntak	Grouping of Business Fields for Productive Micro Business Assistance (BPUM) Based on Deli Serdang Region Using the K-Means Clustering Method (Case	K-Means Clustering Method	This research resulted in 3 groups of MSMEs based on business fields, which were identified through the K-Means Clustering method. Clustering Results:

Researcher/Year	Title	Method	Output
	Study: Dinas Koperasi Dan UMKM Deli Serdang Regency)		- Group 1: MATLAB: Consists of 383 MSME data. RapidMiner: Consists of 371 MSME data. - Group 2: MATLAB: Consists of 261 MSME data. RapidMiner: Consists of 281 MSME data. - Group 3: MATLAB: Consists of 360 MSME data. RapidMiner: Consists of 352 MSME data. Interpretation: Although there is a slight difference in the amount of data in each group between MATLAB and RapidMiner

Researcher/Year	Title	Method	Output
			results, in general, this study shows that the K-Means Clustering method successfully groups MSME data based on business fields.
Haryati Rian Piarna Nunu Nugraha Purnawan Dwi Vernanda Tri Herdiawan Apandi Nurfitria Khoirunnisa	Analysis of Consumer Characteristics and Dominant Factors for Successful Promotion of MSMEs in Subang Regency on social media	K-Means Clustering Analytical Hierarchy Process (AHP Focus Group Discussion (FGD)	This research produced two main outputs: Consumer Clustering (K-Means Clustering): Clustering Results: The K-Means method successfully grouped 70 MSME consumer respondents in Subang Regency into 2 clusters with a DBI (Davies-Bouldin Index) value of 0.089. This DBI value indicates that the separation between the two

Researcher/Year	Title	Method	Output
			clusters is quite good. Cluster 0: Consists of 44 respondents. Cluster 1: Consists of 26 respondents. Interpretation: This study does not provide specific information about the characteristics of consumers in each cluster. To find out the characteristics of consumers in each cluster, access to the full study is required. The full study will provide more detailed information about: The criteria used to categorize the data: For example, age range, gender, social media used, product

Researcher/Year	Title	Method	Output
			category searched, and product criteria. Analysis of the characteristics of each cluster: For example, the dominant type of business in each cluster, average turnover, and so on.
			Dominant Factors of Promotion Success (Analytical Hierarchy Process - AHP): Analysis Results: AHP identified 5 dominant factors that influence the success of MSME promotion on social media: Product Review: Consumers strongly consider product reviews before buying.

Researcher/Year	Title	Method	Output
			Quality of Service to Consumers: Good service quality is an important factor in building consumer trust. Quality Content: Promotional content that is engaging, informative, and relevant to the target market is essential. Influencer Services: Utilizing the services of influencers can increase the reach and credibility of promotions. Actively Engage in Online Community: Actively interacting with consumers on social media can build engagement and loyalty. Interpretation: The results of the AHP analysis show

Researcher/Year	Title	Method	Output
			that factors related to product quality, service, and promotional content have a huge influence on the success of MSME promotion on social media. Utilizing the services of influencers and active interaction on social media is also important to increase promotional reach and engagement.
Hendra Achmadi Danet Patria Dewi S Wuisan Pauline H.P. Tan	Antecedent Intention to Use Mobile Payment at Pasar Rakyat Micro Enterprises in Jakarta During the Covid-19 Pandemic and Clustering the Characteristic	Quantitative Analysis: Partial Least Squares Structural Equation Modeling (PLS-SEM) using SmartPLS software to analyze the relationships	Quantitative Findings: Enthusiasm has a strong positive impact on perceived usefulness (0.899). Anxiety has a weaker positive impact on perceived usefulness (0.18).

Researcher/Year	Title	Method	Output
	Respondent using K-Means	between variables (enthusiasm, anxiety, perceived usefulness, intention to use mobile payments). Data Mining: K-means clustering using R Studio software to identify groups of respondents with similar characteristics based on their profiles (sales volume, education, network provider, number of suppliers, number of customers, gender).	Enthusiasm has a moderate positive effect on intention to use (0.463). Perceived usefulness has a significant positive impact on intention to use (0.312). Anxiety has a negative impact on intention to use (-0.226). Clustering Findings: The study identified clusters of respondents with distinct profiles. One cluster, characterized by lower sales volume, a higher proportion of respondents with an SMA (high school) education, Indosat network usage, a smaller number of suppliers and customers, and

Researcher/Year	Title	Method	Output
			a majority of male respondents, showed a higher intention to use mobile payments.
Rezzy Eko Caraka Robert Kurniawan Prana Ugiana Gio Mohammad Basyuni Bahrul Ilmi Nasution Jamilatuzzahro Jamilatuzzahro Bens Pardamean	Micro, Small, and Medium Enterprises' Business Vulnerability Cluster in Indonesia: An Analysis Using Optimized Fuzzy Geodemographic Clustering	Fuzzy Spatial Clustering: The study employed a fuzzy spatial clustering approach to analyze the business vulnerability of MSMEs in Indonesia. This approach has been previously used to analyze social vulnerability to natural hazards. Optimization with Flower Pollination Algorithm (FPA): The study proposed using the FPA to optimize the Fuzzy Geographically Weighted Clustering (FGWC) method. This optimization	Performance Comparison: The results showed that FPAFGWC outperformed other methods in optimizing FGWC clustering for business vulnerability analysis. Business Vulnerability Clusters: The study identified clusters of business vulnerability across Indonesia. Key Findings: Most regions outside Java Island have vulnerable businesses.

Researcher/Year	Title	Method	Output
		aimed to improve the accuracy and effectiveness of the clustering process. Comparison with other optimization algorithms: The researchers compared the performance of FPAFGWC with traditional FGWC and other optimization algorithms like Artificial Bee Colony, Intelligent Firefly Algorithm, Particle Swarm Optimization, and Gravitational Search Algorithm.	Java Island, particularly the JABODETABEK area (Jakarta, Bogor, Depok, Tangerang, Bekasi), which is the national economic backbone, has generally less vulnerable businesses. Recommendation: The study recommends addressing the gap between the number of MSMEs in Indonesia, potentially through targeted support and intervention strategies based on the identified vulnerability clusters.
Sunday Noya Novenda Kartika Putrianto	Classification of Small and Medium Industry in Malang Regency	K-Means Clustering: The study used the K-means clustering method to classify	Clustering Results: The SMIs were classified into 3 clusters, with patterns based on

Researcher/Year	Title	Method	Output
Weshley Valentino Salendu	Using K-Means based Clustering Method	SMIs based on five variables: Type of business Investment value Business age Product variations Average monthly income	investment value, business age, and average monthly income. Key Finding: Nearly 90% of the SMIs were grouped into clusters with the lowest specifications, indicating a potential need for support and development initiatives to improve the overall condition of SMIs in Malang Regency. This study highlights the importance of data-driven approaches to understand the characteristics and needs of SMIs in a region. By applying K-means clustering, the researchers were able to identify distinct groups of

Researcher/Year	Title	Method	Output
			SMIs based on key factors, providing valuable insights for stakeholders involved in supporting and developing the SMI sector in Malang Regency.
Nazaruddin Malik Idah Zuhroh Muhammad Sri Wahyudi Suliswanto Mochamad Rofik	Sustainable Development Clustering in East Java Using the K-means Method	K-Means Clustering: The researchers used the K-means clustering method to identify clusters of sustainable development in East Java Province. The choice of K-means was based on its accuracy for large datasets.	Clustering Results: The study identified clusters of sustainable development in East Java. Key Findings: Economic performance was found to have a non-significant impact on the occurrence of sustainable development in East Java. Welfare (presumably social and human well-being indicators) was found to have a

Researcher/Year	Title	Method	Output
			significant influence on the clustering of sustainable development in East Java. This study uses data clustering to explore the spatial patterns of sustainable development in East Java. The findings suggest that while economic performance might not be the primary driver of sustainable development in the region, factors related to welfare and social well-being play a crucial role. This information can be valuable for policymakers and stakeholders in East Java to prioritize policies and interventions that promote sustainable development,

Researcher/Year	Title	Method	Output
			particularly in areas with lower welfare indicators.
M G F Christine M Dachyar R Nurcahyo	Product Segmentation of Wooden Handicraft Micro, Small and Medium Enterprises (MSMEs) in Indonesia	Product Segmentation: The study used a combination of K-Means Algorithm and Frequency, Monetary, Customer Variety (FMC) variables to segment the products based on customer transaction data. K-Means Algorithm: This algorithm helps group similar products based on their characteristics. FMC Variables: These variables measure the frequency of purchases, the monetary value of transactions, and the variety of customers	Segmentation Results: The study identified five different types of product sales characteristics in three wooden handicraft MSMEs. Key Finding: The segmentation analysis provides valuable insights into the product sales characteristics, which can be used by business owners to develop targeted strategies for product development and business growth in a competitive market. This study demonstrates how data analysis techniques like K-means clustering

Researcher/Year	Title	Method	Output
		purchasing the products.	and FMC variables can be applied to segment products based on customer behavior. This information helps wooden handicraft MSMEs in Yogyakarta understand their customer base better, tailor their product offerings, and develop more effective business strategies to thrive in a competitive market.

CHAPTER III
RESEARCH METHODOLOGY

3.1 Research Objective

The primary objective of this research is to investigate the challenges and barriers faced by Micro, Small, and Medium Enterprises (MSMEs) in obtaining halal certification in Indonesia. This objective is crucial in understanding the difficulties that MSMEs encounter in the process of obtaining halal certification, which can inform the development of targeted interventions to support their adoption of halal certification.

In addition to the primary objective, this research also aims to examine the current state of halal certification adoption among MSMEs in Indonesia and identify the factors influencing their readiness for certification. This objective seeks to provide a comprehensive understanding of the current landscape of halal certification among MSMEs in Indonesia, including the level of awareness, adoption rates, and factors that influence their decision to obtain halal certification.

Furthermore, this research aims to explore the impact of regulatory requirements on the adoption of halal certification among MSMEs in Indonesia. This objective is essential in understanding how regulatory requirements, such as the implementation of halal certification obligations, affect the adoption of halal certification among MSMEs, and how these requirements can be streamlined to support the growth of the halal industry.

3.2 Research Subject

The subjects of this research are Micro, Small, and Medium Enterprises (MSMEs) in Indonesia. Specifically, the research focuses on MSMEs that are involved in the halal industry or have the potential to enter the halal market in the region of Yogyakarta City.

3.3 Data Sources

The research is supported by two categories of data: primary and secondary data. The primaryresearch data and secondary research data are outlined below.

3.3.1 Primary Data

Primary data in this study were obtained from segmenting the MSMEs using the google form. The direct observation is mandatory to convey the google form from the researcher to the MSMEs owner.

3.3.2 Secondary Data

Secondary data refers to data that has been previously acquired by someone else for a reason other than the current study activity. It is data that already exists and has been gathered by other scholars, organizations, or government bodies.

3.4 Data Collection Method

3.4.1 Literature Study

A iterature study was conducted to obtain relevant literature related to the methods used in this research. Literature sources come from books, journals, and previous studies. The data obtained based on this literature study are inductive studies and deductive studies. In addition, the attributes used for the clustering process were also selected using considerations from inductive studies or previous research on factors that influence community behavior.

The following are some of the variable attributes used to create questionnaire questions:

Table 3.2 Research Literature Study

No	Attributes	Information
1	Demography (Rachman, 2023)	Demographics in this study relate to the characteristics of MSMEs such as district / city, length of business, number of products, age of MSME owners, gender of MSME owners, latest education of MSME owners, and monthly income of MSMEs.
2	Knowledge (Theresia & Rostiana, 2024)	Knowledge of halal certification requirements and procedures is essential for MSMEs to understand what is needed to obtain such certification. This knowledge includes an understanding of halal certification, methods and requirements for applying for halal certification, halal ingredients, and compliance with applicable halal standards and regulations.

No	Attributes	Information
3	Cost (*I Gusti Ayu Mirah Sri Wijayani*, 2019)	Knowledge of the costs of obtaining halal certification is indeed an important attribute in research, because this can influence the decision of MSME actors to process certification. The cost factor can be one of the barriers or motivations for MSMEs.
4	Technology Readiness (Arifah et al., 2022)	Sufficient business technology readiness to be ready to have halal certification. as well as sufficient access to information to access information about procedures for obtaining halal certification. together with the attribute "information access" these two attributes can collaborate with one another.
5	Awareness (Trianda, 2023)	Awareness of the importance of halal certification in a market that is increasingly aware of the halal level of products is an impetus for MSMEs to obtain this certification. This awareness can also increase the competitiveness of MSME halal products in domestic and international markets.
6	Interest (*Muhaimin*, 2022)	Interest from business owners to have halal certification for MSMEs to increase their business credibility.
7	Perception (Ihram, 2020)	Perceptions of MSME owners in the ownership of halal certification, as well as the impact on their MSMEs that already have halal certification.
8	Attitudes (Indraswari et al., 2023)	The attitude of MSMEs owners towards halal certification. Whether or not ownership of halal certification is important, as well as the impact of ownership of halal certification on MSMEs.

3.4.2 Observation

The data collection method using the observation method in the context of MSME halal certification involves direct observation of activities, behaviors, and interactions that occur in the field related to this topic. The observation method can provide valuable insights about the practices and challenges faced by MSMEs related to halal certification, as well as provide a first-hand view of how MSMEs operate in this context.

3.4.3 Questionnaire

The questionnaire is designed using variables derived from literature reviews and will be administered directly to MSME participants in the Yogyakarta region. This data collection employs a purposive sampling technique, targeting MSME actors specifically in Yogyakarta. By utilizing purposive sampling, researchers can choose samples that are highly specific and pertinent to the research goals, facilitating the acquisition of more targeted and comprehensive data. Below is a draft of the questionnaire intended for MSME actors, with the collected data to be analyzed using the Clustering method.

Table 3.3 Questionnaire Question Draft

No	Attributes	Indicator
1	Demography	Length of business Age of MSME Owner Last education of MSME owner MSME monthly income
2	Knowledge	Have you ever attended training? Do you know the requirements for applying for halal certification? Know how to apply for halal certification? Knowing what documents must be owned to apply for halal certification? Know the difficulties in applying for halal certification and how to overcome them?
3	Cost	Do you prioritize increasing business capital over conducting halal certification? Do you face obstacles in budgeting the costs required to carry out halal certification? Do you feel that the cost is a heavy burden on your business? Is your income sufficient to carry out halal certification? Is the cost of halal certification expensive?
4	Information Access	Have you received information about halal certification?

No	Attributes	Indicator
		Is it easy to access information regarding the processing of halal certification? Has there been training related to education and understanding of halal certification? Is it easy to take care of halal certification easy to reach? Is the information regarding halal certification clear?
5	Technology Readiness	Does the owner or staff have computer skills? Does the owner or staff have the ability to access the internet? Are there people around the MSME who can help operate computers and access the internet? Does the managed MSME have a personal computer? Does the MSME have its own internet access?
6	Awareness	Awareness of the benefits of halal certification? Awareness that food and beverage products must be halal certified? Do you guarantee that products that do not have halal certification are safe? Awareness of the importance of halal certification for the community? Do the products you sell have to be halal certified?

3.5 Data Analyzing Method

The data processing method that was utilized in this research was carried out when the data needed was sufficient for the research to be carried out. The first step in the data processing method is to collect data from the owners of MSMEs. This is done by creating a questionnaire using Google Forms, which is then shared with the owners of MSMEs.

The questionnaire should include relevant questions such as the MSME owner's name, MSME name, industry, number of employees, annual revenue, and other relevant questions. Once the questionnaire is shared, the responses are collected and stored for further processing. The next step is to determine the weights of the variables using the Analytical Hierarchy Process (AHP). This is done by creating an Excel questionnaire for pairwise comparison, which is then shared with experts. The experts are asked to fill out the questionnaire, and their responses are collected. A consistency test is then performed to ensure that the data is consistent.

If the data is not consistent, the experts are asked to refill the questionnaire. After the data is collected and the weights are determined, the next step is to preprocess the data. This involves collecting and organizing the data from the questionnaire responses, removing any missing or duplicate values, and handling any outliers. The data is then transformed into a suitable format for clustering. The next step is to perform clustering using the K-means algorithm. This involves determining the optimal number of clusters (K) using the Elbow method or the Silhouette Coefficient.

Once the optimal K value is determined, the K-means algorithm is used to determine the centroid of each cluster. The final step is to cluster the objects based on their closest distances to the centroids. This is done using the K-means algorithm, and the objects are assigned to the cluster with the closest centroid. Additionally, a check is performed to see if any objects are moving between clusters.

3.6 Research Flowchart

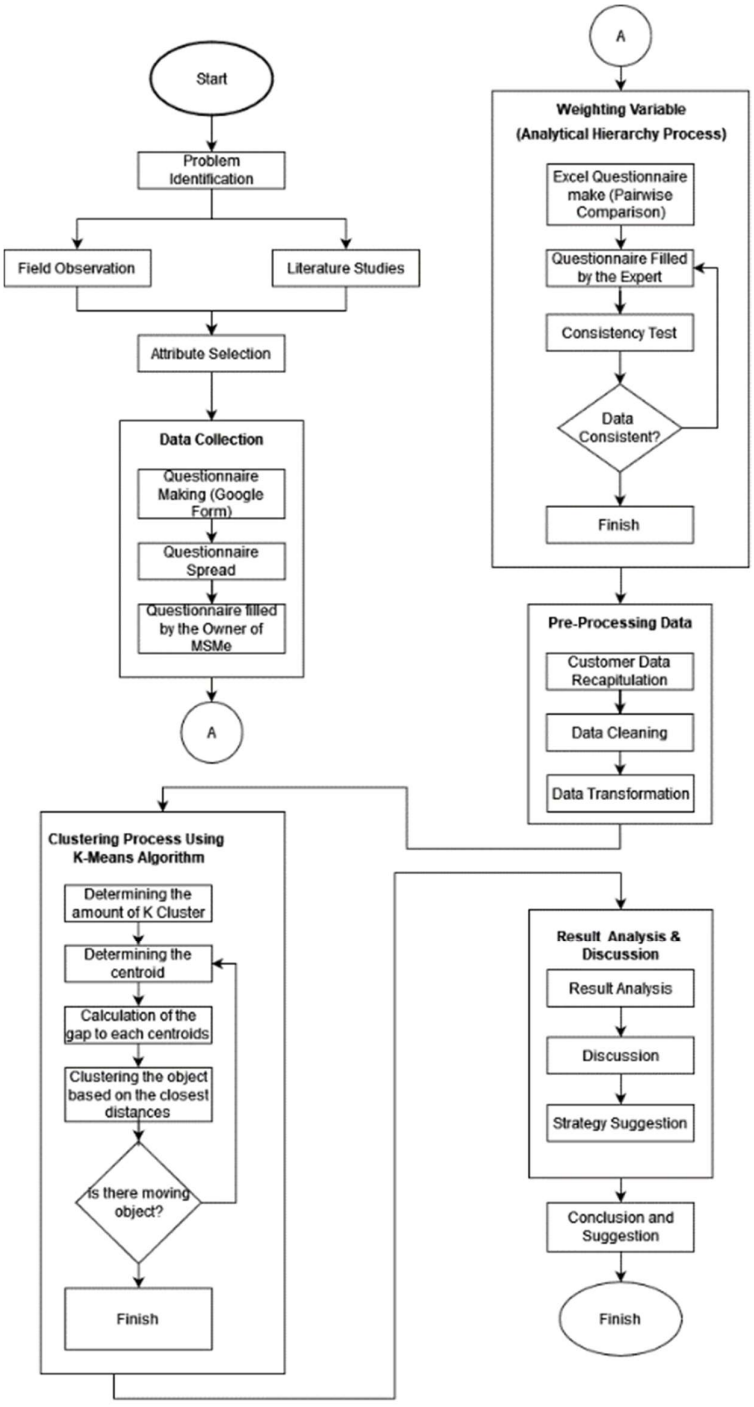


Figure 3.1 Research Flowchart

Figure 3.1 above, contains information about the course of research that will be described in the explanation below:

1. Start, this is the beginning of the flowchart
2. Problem Identification, consists of field observation and literature studies
3. Attribute Selection, Attribute is where we select relevant attributes or variables to focus on.
4. Data Collection, Collect data through a questionnaire (Google Form) distributed to MSME owners.
5. Weighting Variable, Use the Analytical Hierarchy Process (AHP) to assign weights to the selected attributes. The expert will be the one who assigns the weight to each variable used.
6. Pre-processing Data, This step involves cleaning and preparing the data for analysis. This can involve handling missing values, data normalization, and data transformation. This step consists of: Customer data recapitulation, the customer data is recapitulated, which involves summarizing and organizing the data to prepare it for clustering. Data Cleaning This step involves cleaning the data to remove any errors, outliers, or inconsistencies.
7. Clustering Process Using K-Means Algorithm, In this step, the pre-processed data is clustered using the K-means algorithm. Clustering is a technique used to group similar data points into clusters, and K-means is a popular algorithm for doing so. This step involves determining the optimal number of clusters (K) for the data. The centroid of each cluster is calculated, which represents the central point of each cluster.
8. Result Analysis & Discussion, in this step, the clusters obtained from the K-means algorithm are analyzed to identify patterns, trends, and relationships between the variables.
9. Finish

CHAPTER IV
DATA COLLECTION AND PROCESSING OF SYSTEM DEVELOPMENT

4.1 Introduction

Before the research could begin, the first ever step to assess the halal certification readiness of Micro, Small, and Medium Enterprises (MSMEs) in Yogyakarta, a systematic approach to data collection and processing is crucial. This chapter outlines the methodologies employed to gather and analyze data, ensuring that the findings are both valid and reliable. The data collection process is designed to capture a comprehensive picture of the MSMEs' current status regarding halal certification, which is increasingly important in the context of Indonesia's predominantly Muslim population. By utilizing both qualitative and quantitative methods, this research aims to identify the factors influencing readiness for certification and to classify MSMEs based on their preparedness.

4.2 Data Collection

The data collection process is the process of collecting data from various sources to be analyzed and used in a study. Data collection has the aim of gaining a better understanding of the phenomenon being studied, as well as supporting the course of research in order to produce valid research output. This study uses 2 types of data, namely, data from questionnaires conducted by food MSME owners and data from pairwise comparison questionnaires conducted by experts in the field of MSMEs and halal certification. Collecting questionnaire data entitled "Grouping MSMEs Based on Readiness for Halal Certification in the Special Region of Yogyakarta" was carried out through Google Forms media distributed to owners of food MSMEs in Yogyakarta City. In this study, there are 6 main criteria in the questionnaire that were made and distributed to MSMEs in Yogyakarta City, namely, Demographics, Awareness, Knowledge, Costs, Access to information, and Technology Readiness. Demographics are divided into 4 question indicators which are used as criteria, so that for this study there are a total of 9 criteria.

The second dataset is pairwise comparison data filled in by experts through questionnaires. Filling out the questionnaire is done using Microsoft Excel software. The variables used in the pairwise comparison are 9 variables, namely; Awareness, Knowledge, Cost, Information Access, Technology Readiness, MSME Income, Owner's Last Education, Length of Business, and Age of MSME Owner.

4.3 Pre-Processing Data

After the data collection process is completed, the next step is to prepare the data through pre-processing. This stage aims to clean the data that will be used for analysis. Data cleaning is carried out to address issues such as noise, duplication, inconsistency, and incomplete data. As mentioned earlier, 108 data points were collected through online and offline methods over the course of one month. In the process of data summarization and data cleaning, it was found that there were some missing and incomplete data, which was caused by some respondents not filling in or missing, so the data could not be used for the next process. There was also inconsistent data because some respondents answered questions not in accordance with the format provided. Therefore, the amount of data that can be used for analysis is 106.

4.4 Data Transformation

4.4.1 Transformation Rules

The data transformation stage is where a change in the form of the data occurs for the needs of data mining operations. The purpose of data transformation is to improve the quality of the data, make it easier to interpret, or prepare it for use in a particular algorithm. This process may include various steps such as normalization, standardization, clustering, encoding of category variables, removal of duplicates, or handling of missing values. Here are the transformation rules for the existing dataset:

Table 4.4 Variable Attribute Code

Code	Variable Attribute
Z1	Monthly Income of the MSMEs
Z2	Last Education of MSMEs Owner
Z3	Length of Business
Z4	MSMEs Owner Ages

Code	Variable Attribute
Z5	Awareness towards the halal certification
Z6	Knowledge about the halal certification
Z7	Cost
Z8	Information Access about the halal certification
Z9	Technology Readiness

Table 4.5 Transformation Rules

MSMES Owner Ages	< 17 Years old	1
	17 - 25 Years old	2
	25 - 32 Years old	3
	32 - 39 Years old	4
	39 - 46 Years old	5
	>47 Years old	6
Last Education of the MSMES owner	Elementary School (SD)	1
	Junior Highschool (SMP)	2
	Senior Highschool (SMA)	3
	Undergraduate (Sarjana)	4
	Not in School	5
Length of Business	< 1 Years old	1
	1 - 5 Years old	2
	5 - 10 Years old	3
	10 - 15 Years old	4
	15 - 20 Years old	5
	20 - 25 Years old	6
	> 25 Years old	7
Monthly Income of the MSMES	<500.000	1
	500.000 - 5.000.000	2
	5.000.000 - 10.000.000	3
	10.000.000 - 15.000.000	4
	>15.000.000	5
	>25.000.000	6
	>50.000.000	7
Knowledge	Yes	1
	No	2
Cost	Yes	1

	No	2
Information Access about the halal certification	Yes	1
	No	2
Technology Readiness	Yes	1
	No	2
Awareness	Yes	1
	No	2

4.4.2 Analytical Hierarchy Process Data Transformation

The data transformation process for Analytical Hierarchy Process data processing is the process of converting subjective assessments given by experts into numerical values that can be used to determine the relative weight of each criterion in a multicriteria decision.

The following are the transformation process for each expert. 5 experts filled out the questionnaire:

Table 4.6 Expert 1 Pairwise Comparison Matrix

E1

Variable	Z7	Z5	Z6	Z8	Z9	Z4	Z2	Z3	Z1
Z7	1.000	0.333	6.000	2.000	3.000	6.000	5.000	6.000	4.000
Z5	3.000	1.000	3.000	2.000	4.000	7.000	6.000	7.000	5.000
Z6	0.167	0.333	1.000	2.000	2.000	4.000	3.000	5.000	2.000
Z8	0.500	0.500	0.500	1.000	2.000	5.000	3.000	6.000	3.000
Z9	0.333	0.250	0.500	0.500	1.000	3.000	2.000	4.000	2.000
Z4	0.167	0.143	0.250	0.200	0.333	1.000	0.500	2.000	0.333
Z2	0.200	0.167	0.333	0.333	0.500	2.000	1.000	2.000	0.500
Z3	0.167	0.143	0.200	0.167	0.250	0.500	0.500	1.000	0.250
Z1	0.250	0.200	0.500	0.333	0.500	3.000	2.000	4.000	1.000

E1

Variable	Z7	Z5	Z6	Z8	Z9	Z4	Z2	Z3	Z1
Total	1.000	0.333	6.000	2.000	3.000	6.000	5.000	6.000	4.000

Table 4.7 Expert 2 Pairwise Comparison Matrix

E2

Variable	Z7	Z5	Z6	Z8	Z9	Z4	Z2	Z3	Z1
Z7	1.000	0.500	0.500	0.500	2.000	1.500	2.000	2.000	1.500
Z5	2.000	1.000	2.000	4.000	3.000	5.000	4.000	5.000	5.000
Z6	2.000	0.500	1.000	3.000	3.000	4.000	4.000	5.000	6.000
Z8	2.000	0.250	0.330	1.000	2.000	3.000	3.000	4.000	3.000
Z9	0.500	0.330	0.330	0.500	1.000	2.000	3.000	2.000	2.000
Z4	0.670	0.200	0.250	0.330	0.500	1.000	0.500	1.000	0.250
Z2	0.500	0.250	0.250	0.330	0.330	2.000	1.000	2.000	0.500
Z3	0.500	0.250	0.250	0.250	0.330	2.000	0.500	1.000	0.500
Z1	0.670	0.200	0.170	0.330	0.500	3.000	2.000	2.000	1.000
Total	9.84	3.48	5.08	10.24	12.66	23.5	20	24	19.75

Table 4.8 Expert 3 Pairwise Comparison Matrix

E3

Variable	Z7	Z5	Z6	Z8	Z9	Z4	Z2	Z3	Z1
Z7	1.00	0.33	0.25	0.50	2.00	1.50	2.00	2.00	1.50

E3

Variable	Z7	Z5	Z6	Z8	Z9	Z4	Z2	Z3	Z1
Z5	3.00	1.00	2.00	4.00	3.00	5.00	4.00	5.00	6.00
Z6	4.00	0.50	1.00	3.00	3.00	4.00	4.00	5.00	7.00
Z8	2.00	0.25	0.33	1.00	2.00	3.00	3.00	4.00	3.00
Z9	0.50	0.33	0.33	0.50	1.00	2.00	3.00	2.00	2.00
Z4	0.67	0.20	0.25	0.33	0.50	1.00	0.50	1.00	0.25
Z2	0.50	0.25	0.25	0.33	0.33	2.00	1.00	2.00	0.50
Z3	0.50	0.33	0.25	0.25	0.33	2.00	0.50	1.00	0.50
Z1	0.67	0.20	0.17	0.33	0.50	3.00	2.00	2.00	1.00
Total	12.84	3.39	4.83	10.24	12.66	23.50	20.00	24.00	21.75

Table 4.9 Expert 4 Pairwise Comparison Matrix

E4

Criteria	Z7	Z5	Z6	Z8	Z9	Z4	Z2	Z3	Z1
Z7	1.00	0.33	0.33	0.50	3.00	2.00	3.00	3.00	2.00
Z5	3.00	1.00	2.00	5.00	4.00	7.00	6.00	4.00	7.00
Z6	3.00	0.50	1.00	3.00	4.00	5.00	5.00	5.00	8.00
Z8	2.00	0.20	0.33	1.00	2.00	3.00	3.00	4.00	3.00
Z9	0.33	0.25	0.25	0.50	1.00	2.00	4.00	3.00	2.00
Z4	0.50	0.14	0.20	0.33	0.50	1.00	0.33	0.50	0.25
Z2	0.33	0.17	0.20	0.33	0.25	3.00	1.00	2.00	0.50
Z3	0.33	0.25	0.20	0.25	0.33	2.00	0.50	1.00	0.50
Z1	0.50	0.14	0.13	0.33	0.50	4.00	2.00	2.00	1.00
Total	11.00	2.99	4.64	11.25	15.58	29.00	24.83	24.50	24.25

Table 4.10 Expert 5 Pairwise Comparison Matrix

E5

Criteria	Z7	Z5	Z6	Z8	Z9	Z4	Z2	Z3	Z1
Z7	1.00	1.00	1.00	1.00	1.00	0.50	1.00	0.20	1.00
Z5	1.00	1.00	1.00	1.00	1.00	0.50	1.00	0.50	1.00
Z6	1.00	1.00	1.00	1.00	1.00	0.50	1.00	0.50	0.50
Z8	1.00	1.00	1.00	1.00	1.00	0.50	1.00	0.50	0.50
Z9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	1.00
Z4	2.00	2.00	2.00	2.00	1.00	1.00	3.00	3.00	1.00
Z2	1.00	1.00	1.00	1.00	1.00	0.33	1.00	0.20	1.00
Z3	5.00	2.00	2.00	2.00	2.00	0.33	5.00	1.00	1.00
Z1	1.00	1.00	2.00	2.00	1.00	1.00	1.00	1.00	1.00
Total	14.00	11.00	12.00	12.00	10.00	5.67	15.00	7.40	8.00

4.4.3 Clustering Data Transformation

Clustering Data transformation process is a processing step where the answers for each variable are converted from text to nominal form. This stage is carried out to ensure that the data taken is in accordance with the research needs. The following are the results of data transformation for Clustering:

Table 4.11 Transformed Data

No	MSMEs Name	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9
1	Dingo Coffee	5	4	2	3	1.4	2	1.2	2	1.2
2	Soto 40 an	3	3	5	6	1	2	1	1	1.2
3	pecel	1	4	2	2	1	1.2	1	1	1
4	Es Kepal Milo Nikmat Rasa	2	4	2	2	1.2	1.2	1.6	1.2	1.2
5	7Rasa Catering	5	4	3	3	1.4	1.2	1.4	1.8	1.4
6	Funny Warmindo	2	1	2	3	1.4	1.8	1.2	1.6	1.6
7	sate kambing pak udin	2	3	3	5	1	2	1.2	1.8	1.4

No	MSMEs Name	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9
8	WestPash Coffee	2	4	2	2	1	1.4	1.2	1.2	1.4
9	Mie Ayam Afui	5	4	3	5	1	1.6	1.8	1.2	1.2
10	Dawet Ireng Yudhistira	2	4	1	3	1.2	1.2	1.2	1.2	1
…	**…**	**…**	**…**	**…**	**…**	**…**	**…**	**…**	**…**	**…**
...	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
100	gado-gado lotek kupat tahu bu gendut	3	3	2	5	1	1.2	1.6	1	1.4
101	cilok barokah	2	3	2	3	1.4	1.8	1.2	1.8	1.6
102	Bakso tusuk payaman	2	3	1	2	1	2	1.4	1.8	1.4
103	Martabak RR	2	3	2	3	1.8	1.6	1.2	1.8	1.4
104	batagor depan KPRI	2	2	2	4	1	1.8	1.8	1.2	1.4
105	kue pukis	2	1	2	3	1.4	2	1.2	1.8	1.6
106	Bakso akanv emus	3	3	3	4	1	1.6	1.8	1.6	1.2

Table 4. 12 Criterias Description

Name	Description
Z1	Monthly Income of the MSMEs
Z2	Last Education of MSMEs Owner
Z3	Length of Business
Z4	MSMEs Owner Ages
Z5	Awareness about the halal certification
Z6	Knowledge of getting the certification
Z7	Cost
Z8	Information Access about the halal certification
Z9	Technology Readiness

4.5 Analytical Hierarchy Process Data Processing

4.5.1 Arithmetic Means Theory

The arithmetic mean is a measure of concentration that can represent a set of data or describe the center value of a group of numbers. Data centering aims to achieve balance in goal setting while considering other goals simultaneously (*Febrianto Afli*, 2024). The following is the result of a calculation using the Arithmetic Means Theory:

Table 4.13 Arithmetic Means Theory Calculation Result

Criteria	Cost	Awareness	Knowledge	Information Access	Technology Readiness	MSMEs Owner Ages	Last Education of MSMEs Owner	Length of Business	Monthly Income of the MSMEs
Cost	1.00	0.50	1.62	0.90	2.20	2.30	2.60	2.64	2.00
Awareness about the halal certification	2.40	1.00	2.00	3.20	3.00	4.90	4.20	4.30	4.80
Knowledge of getting the Certification	2.03	0.57	1.00	2.40	2.60	3.50	3.40	4.10	4.70
Information Access about the halal certification	1.50	0.44	0.50	1.00	1.80	2.90	2.60	3.70	2.50
Technology Readiness	0.53	0.43	0.48	0.60	1.00	2.00	2.60	2.30	1.80
MSMEs Owner Ages	0.80	0.54	0.59	0.64	0.57	1.00	0.97	1.50	0.42
Last Education of MSMEs Owner	0.51	0.37	0.41	0.47	0.48	1.87	1.00	1.64	0.60
Length of Business	1.30	0.59	0.58	0.58	0.65	1.37	1.40	1.00	0.55
Monthly Income of the MSMEs	0.62	0.34	0.59	0.67	0.60	2.80	1.80	2.20	1.00

4.5.2 Consistency Test

Consistency testing of this data is performed to ensure that the pairwise comparison matrix used in AHP analysis is correct and consistently reliable. Reliability test or consistency test (degree of consistency) is a measure that shows how high an instrument can be trusted or reliable, meaning that reliability concerns the determination of measuring instruments measuring research variables over time (Sari & Arianto, 2020). To conduct a consistency or reliability test, the existing data must be normalized first, so that the range of values between the variables used is normal or uniform. Below is the data that has been normalized:

Table 4.14 Data Normalized

Criteria	Cost	Awareness	Knowledge	Information Access	Technology Readiness	MSMEs Owner Ages	Last Education of MSMEs Owner	Length of Business	Monthly Income of the MSMEs
Cost	0.0935	0.1046	0.2083	0.0861	0.1706	0.1016	0.1264	0.1129	0.1089
Awareness about the halal certification	0.2245	0.2092	0.2577	0.3061	0.2326	0.2165	0.2042	0.1839	0.2613
Knowledge of getting the Certification	0.1902	0.1186	0.1289	0.2296	0.2016	0.1546	0.1653	0.1754	0.2559
Information Access about the halal certification	0.1403	0.0921	0.0642	0.0957	0.1396	0.1281	0.1264	0.1583	0.1361
Technology Readiness	0.0499	0.0904	0.0621	0.0574	0.0775	0.0884	0.1264	0.0984	0.0980
MSMEs Owner Ages	0.0749	0.1124	0.0760	0.0611	0.0439	0.0442	0.0470	0.0642	0.0227
Last Education of MSMEs Owner	0.0474	0.0767	0.0524	0.0445	0.0374	0.0825	0.0486	0.0701	0.0327
Length of Business	0.1216	0.1244	0.0747	0.0558	0.0503	0.0604	0.0681	0.0428	0.0299
Monthly Income of the MSMEs	0.0578	0.0717	0.0756	0.0636	0.0465	0.1237	0.0875	0.0941	0.0544

After normalizing the data as done above, the data is successfully converted into values in each variable that have the same range.

Table 4.15 Consistency Test

Total Weight Matrix	Eugen vector	Multiplication Matrix	Eugen Value	Lambda Max	CI	IR	CR
1.112944	0.12366	1.43817411	11.63002335	11.66840242	0.333550303	1.45	0.230034691
2.096117	0.232902	2.710723189	11.63890747				
1.620002	0.1800	2.104917051	11.69396598				
1.080748	0.120083	1.411454537	11.75398525				
0.748468	0.083163	0.967664884	11.63574716				
0.546374	0.060708	0.703664012	11.59091082				
0.492306	0.054701	0.644673435	11.78548487				
0.627981	0.069776	0.798193521	11.43942823				

Total Weight Matrix	Eugen vector	Multiplication Matrix	Eugen Value	Lambda Max	CI	IR	CR
0.67506	0.075007	0.888617008	11.84716866				
9	1	11.66808175	105.0156218				

Consistency ratio (CR) value is obtained from the division between Consistency Index (CI) and Ratio Index (RI), CI value can be obtained through calculations such as the table above. The IR value is obtained through a random index table determined based on the number of variables used. To determine the level of consistency of AHP method user fields, it must be completed by calculating the Consistency Index. After the consistency index is obtained, the results are compared with the Random Consistency Index (RI) for each n objects (*Rosa de Lima*, 2009). The following is the Random Index Table:

Table 4.16 Ratio Index Table

n	1,2	3	4	5	6	7	8	9	10	11	12	13	14	15
RI	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.48	1.56	1.57	1.59

For this study, 9 criteria variables were used, so the RI value used was 1.45. In the consistency test (Consistency ratio), the consistency ratio (CR) was obtained as 0.0988. Since the CR value of 0.0988 is still below 0.1, your comparison results are considered quite consistent and acceptable. Overall, the CR value of 0.0988 indicates that your pairwise comparison matrix is quite consistent and the calculated weight results are reliable in the research.

4.5.3 Weighting Result

When all values are confirmed to be consistent, the final result is a weight or priority value that shows the relative importance of each criterion or alternative. This result is obtained after calculating the eigenvector of the normalized pairwise comparison matrix, where each weight reflects how important the criterion is in the context of the decision being taken.

Criteria with the highest weight will have the greatest influence, while those with lower weights will be considered less significant. The results of this weighting are then used in the decision-making process or further analysis.

Table 4.17 Criteria Weight Value

Code	Variable	Weight
Z7	Cost	0.117392
Z5	Awareness about the halal certification	0.2415748
Z6	Knowledge of getting the Certification	0.1768163
Z8	Information Access about the halal certification	0.1169408
Z9	Technology Readiness	0.0842214
Z4	MSMEs Owner Ages	0.0652149
Z2	Last Education of MSMEs Owner	0.0543564
Z3	Length of Business	0.0688116
Z1	Monthly Income of the MSMEs	0.0746718

4.6 Clustering Data Processing

4.6.1 Data Standardization

Before starting to process cluster data, it is necessary to consider whether the data units have a significant or large difference. Data standardization is a data processing process that rescales features or variables so that the data has an average (mean) of 0 and a standard deviation of 1 (Gramfort & Michel, 2012). Standardization is important because it helps to scale different data so that machine learning algorithms that are sensitive to data scale, such as distance-based algorithms (e.g., K-Means and K-Nearest Neighbors), can perform more optimally. The following are the data result after the standardization has been done:

- K-means Clustering (Not Weighted)

Table 4.18 Standardize K-Means Clustering

Clustering K-MEANS									
No	**Z1**	**Z2**	**Z3**	**Z4**	**Z5**	**Z6**	**Z7**	**Z8**	**Z9**
1	2.75	1.16	-0.46	-0.98	0.77	0.97	-0.69	1.42	-0.64
2	0.64	0.14	1.95	1.19	-0.80	0.97	-1.42	-1.37	-0.64
3	-1.47	1.16	-0.46	-1.71	-0.80	-1.29	-1.42	-1.37	-1.35
4	-0.42	1.16	-0.46	-1.71	-0.01	-1.29	0.77	-0.82	-0.64
5	2.75	1.16	0.34	-0.98	0.77	-1.29	0.04	0.86	0.07
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…

Clustering K-MEANS									
102	-0.418	0.136	-1.267	-1.707	-0.800	0.974	0.041	0.858	0.067
103	-0.418	0.136	-0.463	-0.983	2.341	-0.161	-0.687	0.858	0.067
104	-0.418	-0.893	-0.463	-0.259	-0.800	0.407	1.499	-0.816	0.067
105	-0.418	-1.922	-0.463	-0.983	0.770	0.974	-0.687	0.858	0.776
106	0.637	0.136	0.341	-0.259	-0.800	-0.161	1.499	0.300	-0.642

- K-means Clustering Weighted

Table 19 Standardize Weighted Clustering K-Means

Clustering K-MEANS (Weighted)									
No	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9
1	0.205	0.063	-0.032	-0.064	0.186	0.172	-0.081	0.166	-0.054
2	0.048	0.007	0.134	0.077	-0.193	0.172	-0.166	-0.161	-0.054
3	-0.110	0.063	-0.032	-0.111	-0.193	-0.229	-0.166	-0.161	-0.114
4	-0.031	0.063	-0.032	-0.111	-0.004	-0.229	0.090	-0.095	-0.054
5	0.205	0.063	0.023	-0.064	0.186	-0.229	0.005	0.100	0.006
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
102	-0.031	0.007	-0.087	-0.111	-0.193	0.172	0.005	0.100	0.006
103	-0.031	0.007	-0.032	-0.064	0.565	-0.028	-0.081	0.100	0.006
104	-0.031	-0.049	-0.032	-0.017	-0.193	0.072	0.176	-0.095	0.006
105	-0.031	-0.104	-0.032	-0.064	0.186	0.172	-0.081	0.100	0.065
106	0.048	0.007	0.023	-0.017	-0.193	-0.028	0.176	0.035	-0.054

4.6.2 Determining the Cluster

Optimal clustering is the process of determining the best or most ideal number of clusters in cluster analysis. This determination is important because determining the right number of clusters can improve the accuracy and interpretation of clustering results. There are several common methods used to determine the optimal number of clusters. The methods used to determine the optimum number of clusters are, Gap Statistic and Calinski-Harabasz, comparing the log of the total WCSS obtained from clustering the actual data with the WCSS of the random data. The optimal cluster is found when the Gap Statistic value is maximum, which indicates that the original data is more structured than the random data.

The Calinski-Harabasz Index or also known as Variance Ratio Criterion (VRC) is one of the methods to determine the optimal number of clusters in cluster analysis. This index measures the ratio between inter-cluster variability and within-cluster variability.

The higher the Calinski-Harabasz value, the better the quality of separation between the clusters formed.

1. Gap Statistic Method
    - K-Means Clustering

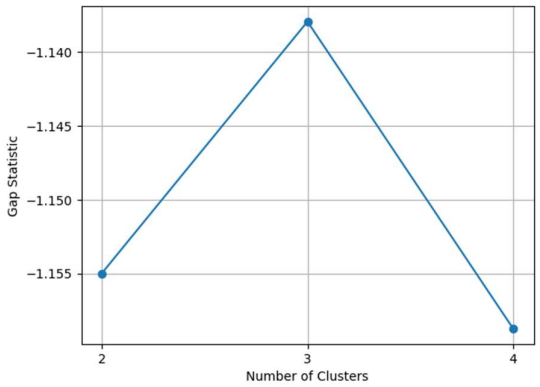


Figure 4.2 Gap Statistic Unweighted

- Weighted K-Means Clustering

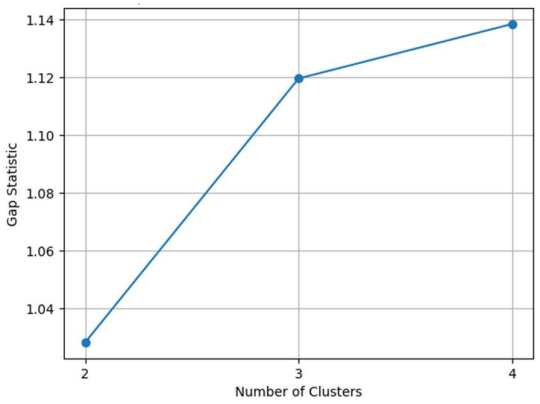


Figure 4.3 Gap Statistic Weighted

Subsequently, clustering was performed with a range of 2 to 5 iterations. Based on the two Gap Statistic graphs above, the results obtained from both methods show different optimal cluster numbers. The unweighted Gap Statistic resulted in 3 clusters as the optimal number, whereas the weighted Gap Statistic resulted in 4 clusters as the optimal number. Therefore, based on the weighted Gap Statistic, 4 clusters are considered the optimal number of clusters. Next, clustering was performed again using a different method, namely Calinski-Harabasz, to confirm the results of the previous clustering.

2. Callinski Harabasz
    - K-Means Clustering

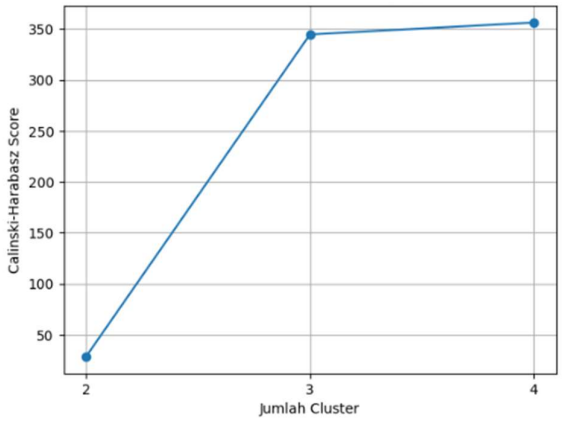


Figure 4.4 Unweighted Calinski Harabasz

- Weighted K-Means Clustering

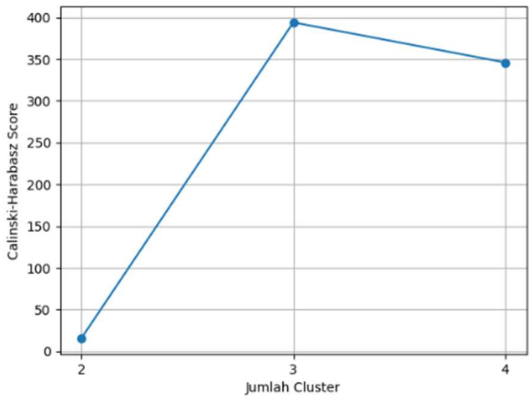


Figure 4.5 Weighted Calinski Harabasz

In the interpretation of Calinski-Harabasz scores, these results provide valuable insights:

The difference in the number of clusters obtained from Gap Statistic and Calinski-Harabasz can be explained by the different ways each method evaluates clustering quality. Gap Statistic measures how well the resulting clustering compares to random data by calculating the difference between the actual and random clustering values. In this case, the unweighted Gap Statistic results in 3 clusters, while the weighted Gap Statistic results in 4 clusters. This difference occurs because the weighting of attributes in Gap Statistic takes into account the relative importance of each feature, which can influence the optimal number of clusters. Meanwhile, the Calinski-Harabasz method evaluates clustering quality differently by assessing how well-separated the clusters are and how low their internal variance is. As a result, Calinski-Harabasz indicates that the data structure is best represented by 2 clusters.

While Gap Statistic suggests more clusters, Calinski-Harabasz sees only two main groups that are well-separated and have higher density, making them more suitable for representing the data. This difference reflects that each method has a different approach to assessing clustering structure, which can lead to different numbers of clusters.

4.6.3 Optimal Cluster Analysis with a Hybrid Approach

Through a hybrid approach, this research utilizes the advantages of each method, where 5 clusters are used for detailed segmentation analysis, while 3 clusters are used on the weighted dataset to obtain a more robust and focused cluster structure. This approach provides flexibility in interpreting the data and achieves a balance between detailed segmentation and optimal cluster compactness. This analysis aims to group data into several clusters that have similar characteristics, thus providing deeper insight into the data structure. provide a deeper insight into the structure of the data. At this stage, clustering analysis is carried out using the optimal value of the number of clusters that has been determined previously.

Clustering is done using the gap statistic method, using a formula using the Python programming language. In Python, using Gap Statistic and Calinski-Harabasz to determine the optimal number of clusters in K-Means clustering requires some additional steps as Gap Statistic is not available as a built-in method in sklearn. Sklearn is the common abbreviation for Scikit-Learn, a popular Python library used for machine learning. When clustering is performed using the K-Means algorithm through a library like sklearn, the results automatically use the Euclidean Distance method to calculate the distance between the data and the centroid. In other words, Euclidean Distance is the default method used by the K-Means implementation in sklearn, so we don't need to calculate the distance manually.

Euclidean Distance is a commonly used method to measure the distance between two points in n-dimensional space. Euclidean distance can be defined as the length of a straight line between two points.

Mathematically, the Euclidean Distance formula between two points P and Q in n-dimensional space.

$$Euclidean\ Distance =$$

$$d(P,Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2}$$

Where:

- p_1 is the n-component from the data point
- q_1 is the n-component from the data point
- n is the number of dimensions of the data (for example, 2 for two dimensions, 3 for three dimensions, and so on.)

After that, each data point is assigned to the cluster with the closest centroid (the one with the shortest distance). Once all data points are assigned to their respective clusters, calculate a new centroid for each cluster. The new centroid is calculated by averaging all the points in the cluster.

$$C_j = \frac{1}{N_j} \sum_{i=1}^{N_j} x_i$$

Where:

- C_j is the new centroid for the j cluster.
- N_j is the total data point inside j cluster.
- x_i is the data point of i in j cluster.

In Python, centroid updates are performed automatically by the K-Means algorithm of the *sklearn cluster KMeans* library. The K-Means algorithm in scikit-learn automatically performs the following steps in each iteration until a convergence condition is reached:

1. Calculates the Euclidean distance between each data point and the centroid.
2. Cluster each data point to the cluster with the closest centroid.
3. Updates the centroid by calculating the average position of all data points in each cluster.

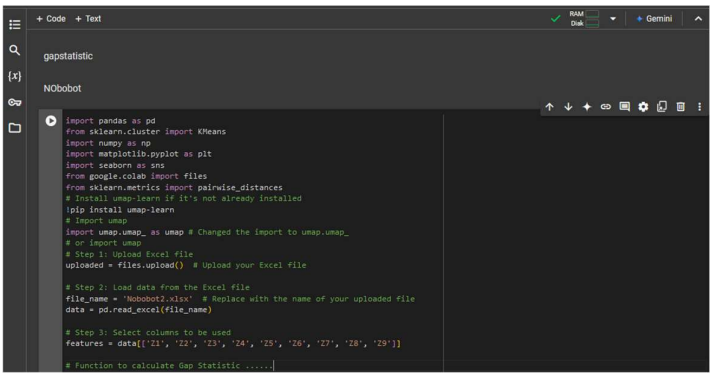


Figure 4.6 Python Using Google Colab

In this study, the difference in the number of clusters obtained between Gap Statistic and Calinski-Harabasz reflects the distinct approaches each method takes in evaluating clustering quality. Gap Statistic shows varying numbers of clusters, both for weighted and unweighted data, because it compares the actual clustering results with those derived from random data. This comparison ensures that the clustering outcome is superior to random clustering. As a result, Gap Statistic is more sensitive to the structure of the data and can yield different numbers of clusters, depending on how the involved attributes or variables influence cluster formation.

On the other hand, Calinski-Harabasz produces consistent results between weighted and unweighted clustering, as this method primarily focuses on the separation between clusters and their internal variance.

Although this method can demonstrate clear separation of clusters, it tends to provide more stable and uniform results, even when the data is weighted. This suggests that, in the context of this research, the separation between clusters identified by Calinski-Harabasz is strong enough to produce similar cluster numbers, regardless of the weighting of the attributes.

Thus, while Gap Statistic results in varying numbers of clusters, it offers the advantage of providing a more comprehensive view of clustering quality by comparing it to random distributions. In contrast, Calinski-Harabasz may be more suitable for data with clear cluster separation and tends to produce stable results despite variations in attribute weights. In this study, the Gap Statistic was used to determine the optimal number of clusters by comparing actual clustering results with those derived from random data. This method ensures that the identified clusters are more structured than random groupings, making it highly sensitive to data structure. The Gap Statistic often produces varying numbers of clusters, depending on how different attributes influence cluster formation, both for weighted and unweighted data. This variability allows for a more comprehensive assessment of clustering quality.

Although other methods, such as the Calinski-Harabasz Index, focus on evaluating cluster separation and internal variance, they tend to produce stable and uniform results regardless of attribute weighting. However, this study prioritizes the Gap Statistic due to its ability to adapt to different data structures and provide a deeper understanding of clustering quality based on randomness comparisons.

The following is the result of clustering based on the Gap statistic that has been calculated:

- K- Means Gap Statistic Unweighted

Table 20 Cluster Member of Unweighted K-Means Clustering

Cluster	Number of Members
1	32
2	39
3	35

- K – Means Gap Statistic Weighted

Table 21 Cluster Member of Weighted K-Means Clustering

Cluster	Number of Members
1	44
2	24
3	10
4	28

4.6.4 Cluster Visualization

Cluster visualization is the process of displaying clustering results in the form of graphs or diagrams, making the patterns and structures within the data more clearly visible. This visualization helps to understand how data is grouped into clusters, the degree of separation between clusters, and the specific characteristics of each cluster. In this study, UMAP (Uniform Manifold Approximation and Projection) is used as a visualization technique for the K-means clustering results. UMAP is a dimensionality reduction technique that projects data from higher dimensions to lower dimensions (typically two or three dimensions) while preserving the topological structure or relationships between data points. UMAP has been increasing in popularity due to their effectiveness in data visualization and finding better representations for downstream tasks such as classification and clustering (Dalmia & Sia, 2021).

- K-means Gap Statistic Unweighted

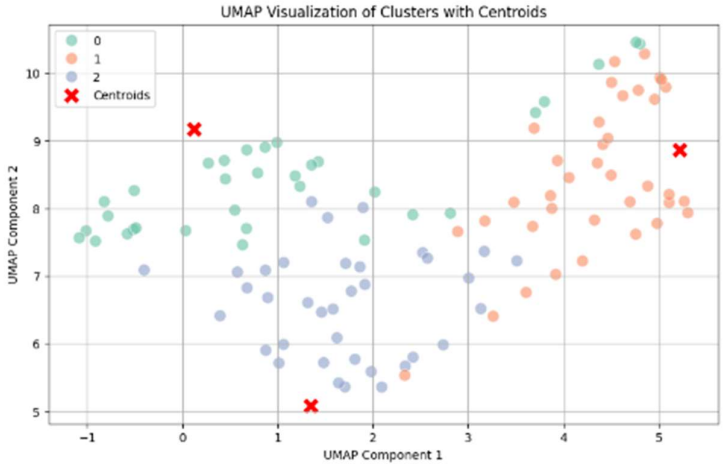


Figure 4.7 UMAP Visualization for Unweighted K-means

- K-means Gap Statistic Weighted

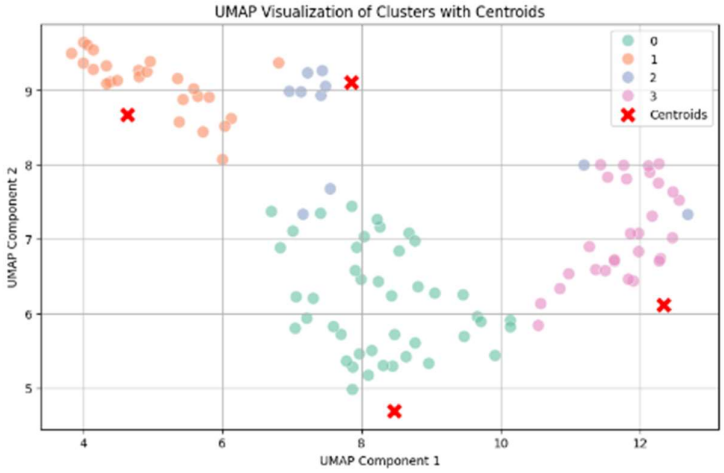


Figure 4.8 UMAP Visualization for Weighted K-means

4.6.5 K-Centroid Value and Final Cluster

The initialization of the data cluster centers where the initial centers are used as the number of classes that will be created is called a k-centroid (Adistya et al., 2023). The centroids are obtained randomly from the N data sets. In the context of the K-Means Clustering algorithm, the centroid is a point that represents the average position of all data belonging to a cluster. The following are the centroids findings:

- K-means Gap Statistic Unweighted

Table 22 Centroid Value of Unweighted K-means

Z-Variable	Value (Cluster 1)	Value (Cluster 2)	Value (Cluster 3)
Z1	-0.41775	-0.24678	0.636572
Z2	0.697034	-0.80949	0.193032
Z3	-0.73087	0.406615	0.252053
Z4	-0.89545	0.483835	0.323551
Z5	-0.13377	0.600551	-0.49461
Z6	0.01135	0.68258	-0.71195
Z7	-0.025	-0.51015	0.547236
Z8	0.063338	0.76767	-0.84705
Z9	-0.25527	0.833003	-0.62214

- K-means Gap Statistic Weighted

Table 23 Centroid Value of Weighted K-Means

Z-Variable	Value (Cluster 1)	Value (Cluster 2)	Value (Cluster 3)	Value (Cluster 4)
Z1	-0.0013	-0.00364	0.003356	0.003967
Z2	0.000582	-0.00447	0.001395	0.003155
Z3	-0.00075	0.000145	-0.0038	0.003585
Z4	0.001886	-0.00182	0.002875	-0.002939
Z5	-0.02676	0.074332	-0.03878	-0.014283
Z6	0.021409	0.026694	-0.02488	-0.048235
Z7	-0.00186	-0.01238	0.035753	-0.006887
Z8	0.005297	0.022749	-0.02018	-0.02075

Z-Variable	Value (Cluster 1)	Value (Cluster 2)	Value (Cluster 3)	Value (Cluster 4)
Z9	0.002318	0.00765	-0.00845	-0.006737

After the centroid value from the K-means clustering result, the next crucial step is to convert the value into the original scale of the data. Because the results of this centroid still cannot be read clearly about the existing cluster groups. The process of converting the centroid values to the original scale produces what is known as the final cluster. The final cluster center is used to determine the attributes that become the dominant assessment of each cluster based on the final cluster center table and the number of cases in each cluster. If the attribute has the highest value in one cluster, then the cluster has a dominant assessment (Asyari & Zagladi, 2020).

- K-means Gap Statistic Unweighted

Table 24 Gap Statistic Unweighted

Z-Variable	Value (Cluster 1)	Value (Cluster 2)	Value (Cluster 3)
Z1	2.00000	2.16216	3.00000
Z2	3.54545	2.08108	3.05556
Z3	1.66667	3.08108	2.88889
Z4	3.12121	5.02703	4.80556
Z5	1.16970	1.35676	1.07778
Z6	1.66061	1.89730	1.40556
Z7	1.38182	1.24865	1.53889
Z8	1.51515	1.76757	1.18889
Z9	1.30909	1.61622	1.20556

- K-means Gap Statistic Weighted

Table 25 Gap Statistic Weighted

Z-Variable	Value (Cluster 1)	Value (Cluster 2)	Value (Cluster 3)	Value (Cluster 4)
Z1	2.3910	3.3726	2.4488	2.4066
Z2	2.8738	3.8403	1.8860	2.8770

Z-Variable	Value (Cluster 1)	Value (Cluster 2)	Value (Cluster 3)	Value (Cluster 4)
Z3	2.5679	1.5744	2.6116	2.5752
Z4	4.3730	3.3429	4.3612	4.3491
Z5	1.1680	1.2956	1.2319	1.1666
Z6	1.6918	1.7074	1.6183	1.4708
Z7	1.3949	1.3722	1.3759	1.3984
Z8	1.4947	1.5331	1.4980	1.4507
Z9	1.3823	1.3942	1.3676	1.3725

4.6.6 Cluster Profile Group

The purpose of cluster profiling is to provide an in-depth understanding of the unique characteristics of each cluster generated in the clustering process. By understanding the profile of each cluster, we can meaningfully interpret the clustering results and use them for strategic decision-making. This process is very important in the interpretation of the clustering results as it provides insight into the unique identity or pattern of each group.

Table 26 Rounding Value of Unweighted K Means

Z-Variable	Value (Cluster 1)	Value (Cluster 2)	Value (Cluster 3)
Z1	3	2	2
Z2	4	2	3
Z3	2	3	3
Z4	3	5	5
Z5	1	1	1
Z6	2	2	1
Z7	1	1	2
Z8	2	2	1
Z9	1	2	1

Table 27 Transformed Rounding Value of Unweighted K Means

Z-Variable	Value (Cluster 1)	Value (Cluster 2)	Value (Cluster 3)
Z1	5000.000 - 10.000.000	500.000 - 5.000.000	500.000 - 5.000.000
Z2	Undergraduate	Junior High	Senior High

Z-Variable	Value (Cluster 1)	Value (Cluster 2)	Value (Cluster 3)
Z3	1 - 5 Years	5 - 10 Years	5 - 10 Years
Z4	25 - 32 Years	39 - 46 Years	39 - 46 Years
Z5	Yes	Yes	Yes
Z6	No	No	Yes
Z7	Yes	Yes	No
Z8	No	No	Yes
Z9	Yes	No	Yes

Table 28 Rounding Value of Weighted K Means

Z-Variable	Value (Cluster 1)	Value (Cluster 2)	Value (Cluster 3)	Value (Cluster 4)
Z1	2	3	2	2
Z2	3	4	2	3
Z3	3	2	3	3
Z4	4	3	4	4
Z5	1	1	1	1
Z6	2	2	2	1
Z7	1	1	1	1
Z8	1	2	1	1
Z9	1	1	1	1

Table 29 Transformed Rounding Value of Weighted K Means

Z-Variable	Value (Cluster 1)	Value (Cluster 2)	Value (Cluster 3)	Value (Cluster 4)
Z1	500.000 - 5.000.000	5000.000 - 10.000.000	500.000 - 5.000.000	500.000 - 5.000.000
Z2	Senior High	Undergraduate	Junior High	Senior High
Z3	5 - 10 Years	1-5 Years	5 - 10 Years	5 - 10 Years
Z4	32 - 39 Years	25 -32 Years	32 - 39 Years	32 - 39 Years
Z5	Yes	Yes	Yes	Yes
Z6	No	No	No	Yes
Z7	Yes	Yes	Yes	Yes
Z8	Yes	No	Yes	Yes
Z9	Yes	Yes	Yes	Yes

Cluster profiling is an analytical process to understand and describe the characteristics of data groups resulting from clustering techniques. Clustering is a method of dividing data into groups based on similarity or common characteristics.

4.7 Comparison Between the Weighted and Unweighted K-Means Clustering

4.7.1 Dataset

After clustering and generating cluster results, the next step is to profile the cluster to find out the group division of the dataset, both for unweighted and weighted. Furthermore, the step taken after that is to compare the results of weighted clusters and unweighted clusters with the aim of seeing if there is a difference between datasets that are weighted with those that are not. The following are the data set for the unweighted and weighted clusters:

- Unweighted K-Means Clustering

Table 30 Unweighted K-Means Clustering Result

No	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9
1	5	4	2	3	1.4	2	1.2	2	1.2
2	3	3	5	6	1	2	1	1	1.2
3	1	4	2	2	1	1.2	1	1	1
4	2	4	2	2	1.2	1.2	1.6	1.2	1.2
5	5	4	3	3	1.4	1.2	1.4	1.8	1.4
6	2	1	2	3	1.4	1.8	1.2	1.6	1.6
7	2	3	3	5	1	2	1.2	1.8	1.4
8	2	4	2	2	1	1.4	1.2	1.2	1.4
9	5	4	3	5	1	1.6	1.8	1.2	1.2
10	2	4	1	3	1.2	1.2	1.2	1.2	1
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
96	2	5	2	5	1	1.8	1.4	1.6	1.4
97	2	3	1	2	1.6	2	1.4	2	1.4
98	2	3	4	6	1	1	1.2	1.2	1.4
99	3	3	2	5	1	1.2	1.6	1	1.4
100	2	3	2	3	1.4	1.8	1.2	1.8	1.6
101	2	3	1	2	1	2	1.4	1.8	1.4
102	2	3	2	3	1.8	1.6	1.2	1.8	1.4
103	2	2	2	4	1	1.8	1.8	1.2	1.4
104	2	1	2	3	1.4	2	1.2	1.8	1.6
105	3	3	3	4	1	1.6	1.8	1.6	1.2

- Weighted K-Mean Clustering

Table 31 Weighted K-Means Clustering

No	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9
1	3	3	3	4	1	2	1	2	1
2	2	3	3	4	1	2	1	1	1
3	2	3	3	4	1	2	1	1	1
4	2	3	3	4	1	2	1	1	1
5	3	3	3	4	1	2	1	2	1
6	2	3	3	4	1	2	1	2	1
7	2	3	3	4	1	2	1	2	1
8	2	3	3	4	1	2	1	1	1
9	3	3	3	4	1	2	1	1	1
10	2	3	2	4	1	2	1	1	1
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
97	2	3	3	4	1	2	1	2	1
98	2	3	2	4	1	2	1	2	1
99	2	3	3	4	1	2	1	1	1
100	2	3	3	4	1	2	1	1	1
101	2	3	3	4	1	2	1	2	1
102	2	3	2	4	1	2	1	2	1
103	2	3	3	4	1	2	1	2	1
104	2	3	3	4	1	2	1	1	1
105	2	3	3	4	1	2	1	2	1
106	2	3	3	4	1	2	1	2	1

4.7.2 T – test Calculation Result

Z1-Z9 represents the unweighted dataset and X1-X9 represents the weighted dataset, so the results of the Paired Samples T-Test analysis and K-Means clustering show a significant effect of weighting on the data structure. Based on the T-Test results, most of the variable pairs between the unweighted and weighted datasets showed statistically significant differences ($p < 0.05$). This confirms that the application of weights changes the average value of the dataset substantially. This change also affected the distribution of the data, as reflected in the K-Means clustering analysis, where the unweighted dataset resulted in 3 clusters, while the weighted dataset resulted in 4 clusters.

Table 32 T-test Calculation

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Z1 - X1	0.29	0.73	0.07	0.15	0.43	4.13	105.00	0.00
Pair 2	Z2 - X2	-0.13	0.98	0.09	-0.32	0.06	-1.39	105.00	0.17
Pair 3	Z3 - X3	-0.31	1.14	0.11	-0.53	-0.09	-2.81	105.00	0.01
Pair 4	Z4 - X4	0.36	1.39	0.13	0.09	0.63	2.66	105.00	0.01
Pair 5	Z5 - X5	0.14	0.35	0.03	0.07	0.21	4.16	105.00	0.00
Pair 6	Z6 - X6	-0.33	0.47	0.05	-0.42	-0.24	-7.19	105.00	0.00
Pair 7	Z7 - X7	0.33	0.47	0.05	0.24	0.42	7.19	105.00	0.00
Pair 9	Z9 - X9	0.26	0.44	0.04	0.18	0.35	6.14	105.00	0.00

Paired Samples T-Test results between the unweighted (Z1-Z9) and weighted (X1-X9) datasets show that weighting has a significant impact on some variables, but not on all. Most pairs of variables (Z1 - X1, Z3 - X3, Z4 - X4, Z5 - X5, Z6 - X6, Z7 - X7, and Z9 - X9) showed statistically significant differences ($p < 0.05$), meaning that the weights affected the mean values of the variables. This indicates that the change in data distribution due to weighting causes a considerable difference between the unweighted and weighted datasets. For example, in Pair 1 (Z1 - X1), the mean value of the difference is 0.29, which means that the weighted dataset has a higher value compared to the unweighted dataset. Conversely, in Pair 6 (Z6 - X6), the average difference is -0.33, indicating that the value on the weighted dataset is lower than the unweighted dataset.

However, there is one pair, Z2 - X2, that does not show a significant difference ($p = 0.17$), which means that weighting does not have a significant effect on this variable. This indicates that not all variables in the dataset undergo substantial changes after weighting.

In this context, the significant difference in most pairs of variables may explain why the K-Means clustering results on the weighted dataset resulted in 4 clusters, while the unweighted dataset resulted in only 3 clusters. Weighting causes changes to the distribution of the data, so the clustering algorithm detects new patterns that are not visible in the unweighted dataset. This difference in the number of clusters indicates that the weighted data is more structured or influenced by certain weights, leading to the formation of additional clusters.

CHAPTER V
RESULT AND DISCUSSION

5.1 The Comparison Between the Normal K-Means and the Weighted K-Means

The comparison between normal K-Means clustering and weighted K-Means clustering in this study highlights the significant impact that weighting has on both the structure and interpretation of the data. When applying normal K-Means to the unweighted dataset, the algorithm identified 3 clusters, which reflected the natural grouping based solely on the inherent patterns and proximities within the raw data. In contrast, applying weighted K-Means to the weighted dataset resulted in 4 clusters, demonstrating that weighting alters the dataset's structure by giving more influence on certain data points. This additional cluster suggests that weighting emphasizes specific aspects of the data, leading to the identification of more distinct or nuanced groupings that were not apparent in the unweighted analysis.

Weighting introduces a bias that prioritizes particular data points, causing the algorithm to adjust the centroids more significantly towards these weighted points. As a result, the boundaries of the clusters are reshaped, allowing the formation of new clusters or the refinement of existing ones. This explains the emergence of an additional cluster in the weighted K-Means analysis, which may represent a subgroup within the data that holds greater significance due to the applied weights. Consequently, the normal K-Means provides an unbiased view of the dataset, suitable for analyses where all data points are considered equally important. On the other hand, weighted K-Means offers a tailored clustering approach that aligns with specific research objectives, such as emphasizing certain demographic groups or prioritizing particular attributes in market segmentation studies.

However, while weighted K-Means can reveal deeper insights by highlighting important subgroups, it also requires careful consideration in the selection of weights to avoid introducing unintended biases.

The choice between normal and weighted K-Means should therefore be guided by the research question at hand. If the goal is to explore the raw structure of the data without any preconceived emphasis, normal K-Means is the appropriate method. Conversely, if the research aims to reflect real-world priorities or emphasize specific aspects of the data, weighted K-Means becomes more suitable. In conclusion, the additional cluster identified through weighted K-Means underscores the importance of selecting the appropriate clustering method based on the dataset's context and the specific objectives of the analysis. This ensures that the clustering results are both meaningful and aligned with the intended focus of the research.

5.2 K-Means Weighted Result

The data provided represents various factors that may influence the performance or success of Micro, Small, and Medium Enterprises (MSMEs), each associated with a specific numerical value that indicates its impact. For instance, the variable "Awareness" (Z5) has the highest value at 0.242, suggesting that the level of awareness among MSME owners has a significant effect on their business outcomes. Similarly, "Knowledge" (Z6) with a value of 0.177 also plays a notable role, implying that the owner's knowledge of business practices or market trends contributes to success. Other factors like "Cost" (Z7) and "Information Access" (Z8) both have a moderate influence with values of 0.117, indicating that the financial and informational resources available to MSMEs are crucial, though not as impactful as awareness or knowledge.

On the other hand, "Technology Readiness" (Z9) and "Length of Business" (Z3) have relatively smaller values of 0.084 and 0.069, respectively, showing that while they are important, their influence is less pronounced. Factors such as "MSMEs Owner Ages" (Z4) and "Monthly Income of the MSMEs" (Z1) also play a smaller role in determining business outcomes, with values of 0.065 and 0.075.

The variable "Last Education of MSMEs Owner" (Z2) has the smallest value of 0.054, suggesting that the level of formal education of the owner has the least impact among the variables listed. Overall, this data highlights that awareness, knowledge, and cost are the most influential factors for MSMEs, while education level and age of the owner have relatively minor effects.

Table 33 Weight on Attributes

Code	Variable	
Z7	Cost	0.117
Z5	Awareness about the halal certification	0.242
Z6	Knowledge of getting the Certification	0.177
Z8	Information Access about the halal certification	0.117
Z9	Technology Readiness	0.084
Z4	MSMEs Owner Ages	0.065
Z2	Last Education of MSMEs Owner	0.054
Z3	Length of Business	0.069
Z1	Monthly Income of the MSMEs	0.075

The prioritization of criteria in the weighted K-Means method better captures the complexities and challenges faced by MSMEs, such as access to information and technological readiness. By emphasizing key factors, the weighted approach aligns more closely with the research objectives and provides targeted insights, making the results more relevant and actionable for understanding and supporting MSMEs.

5.3 Cluster Profiling

Cluster Profiling refers to the process of analyzing and interpreting the characteristics of each cluster obtained after applying a clustering algorithm, such as K-Means, and Hierarchical Clustering. Below is the result of the cluster profiling:

Table 34 Weighted Cluster Profiling Result

Attribute	Cluster			
	1	**2**	**3**	**4**
Z1	500.000 - 5.000.000	5000.000 - 10.000.000	500.000 - 5.000.000	500.000 - 5.000.000
Z2	Senior High	Undergraduate	Junior High	Senior High
Z3	5 - 10 Years	1-5 Years	5 - 10 Years	5 - 10 Years
Z4	32 - 39 Years	25 -32 Years	32 - 39 Years	32 - 39 Years
Z5	Yes	Yes	Yes	Yes
Z6	No	No	No	Yes
Z7	Yes	Yes	Yes	Yes
Z8	Yes	No	Yes	Yes
Z9	Yes	Yes	Yes	Yes

Table 35 Cluster Members

	Members	
Cluster	Unweight	Weight
1	32	44
2	39	24
3	35	10
4	0	28
Total	106	106

The clustering analysis highlights four distinct groups of respondents, each characterized by unique attributes. Cluster 1, identified as "Mid-Income Established Entrepreneurs," consists of individuals with monthly incomes ranging from 500K to 5M. They are predominantly educated at the Senior High level, aged between 32 and 39 years, and have been running their businesses for 5-10 years.

This group demonstrates strong awareness and readiness to adopt technology but faces challenges with knowledge access, suggesting a need for enhanced education or training programs to fill this gap. Cluster 2, termed "High-Income Emerging Entrepreneurs," includes respondents earning between 5M and 10M monthly.

This younger group, aged 25-32 years, has relatively high educational attainment, predominantly at the Undergraduate level, but their businesses are relatively new, operating for only 1-5 years. They show strong awareness and technology readiness but struggle with accessing critical information, indicating a need for resources or tools that improve information dissemination and access.

The monthly income of MSMEs in Cluster 1 ranges from 500,000 to 5,000,000, indicating that this cluster includes businesses with moderate to high income. In Cluster 2, the income range is higher, from 5,000,000 to 10,000,000, suggesting that this group consists of larger, more successful businesses. In Cluster 3, the income range is similar to Cluster 1, between 500,000 and 5,000,000, representing medium-sized enterprises. Cluster 4 shows an income range of 500,000 to 5,000,000, placing it in a similar category to Clusters 1 and 3 in terms of income. In Cluster 1, most MSME owners have a Senior High education, reflecting a relatively high level of education compared to other clusters. Cluster 2 predominantly features owners with an Undergraduate education, indicating a more educated group, which may correlate with better business strategies. In Cluster 3, the owners have a Junior High education, suggesting that they may have less formal education, potentially impacting their business decision-making capabilities. Cluster 4 has a mix of Senior High and Undergraduate education, indicating a balanced educational background among the MSME owners.

Businesses in Cluster 1 have been in operation for 5 to 10 years, suggesting that they are established and have gained significant experience. Similarly, businesses in Cluster 3 also operate within the 5 to 10-year range, indicating maturity and experience. In Cluster 2, the businesses have been operating for 1 to 5 years, indicating a newer and potentially less experienced group. Cluster 4 also shows a length of business between 5 to 10 years, suggesting these businesses have a moderate level of experience, similar to Clusters 1 and 3. In Cluster 1, MSME owners are generally aged 32 to 39 years old, suggesting that this group includes mature and experienced business owners. The owners in Cluster 2 are younger, aged 25 to 32 years, which may imply a more dynamic, innovative approach to business.

In Clusters 1 and 3, the owners are also in the 32 to 39-year age range, indicating a demographic of mature business owners. Cluster 4, similar to Clusters 1 and 3, consists of MSME owners aged 32 to 39 years, highlighting a pattern of experienced business owners across these groups. Awareness is high in all clusters. In Cluster 1, MSME owners have high awareness, indicating they are well-informed about industry challenges and trends. This pattern is consistent across all clusters, as owners in Clusters 2, 3, and 4 also demonstrate high awareness. This suggests that awareness is generally a key characteristic across MSMEs, regardless of other factors.

In Cluster 1, MSMEs show No knowledge, which may indicate a lack of formal or applied business knowledge despite being aware of industry trends. Similarly, MSMEs in Cluster 2 also exhibit No knowledge, suggesting that they may face challenges in applying business insights effectively. In Cluster 3, the lack of knowledge (No) could represent a barrier to utilizing market insights for business growth. In contrast, MSMEs in Cluster 4 exhibit Yes for knowledge, suggesting that they are better equipped with the necessary business knowledge, which may contribute to their success. Cost is a significant factor in all clusters. In Cluster 1, businesses face cost challenges, as indicated by the Yes response. This is consistent with Cluster 2, where MSMEs also prioritize or struggle with managing costs. Similarly, in Cluster 3, cost management is an important concern. In Cluster 4, businesses also indicate a focus on managing costs, suggesting that financial management is crucial across all MSMEs in this study.

Information access is varied across the clusters. In Cluster 1, MSMEs have **Yes** for information access, indicating they have sufficient access to relevant business resources and data. However, businesses in Cluster 2 show **No** for information access, suggesting they may lack the necessary resources or data to make informed decisions. In Cluster 3, businesses again have **Yes** for information access, allowing them to make data-driven decisions. Similarly, Cluster 4 also shows **Yes** for information access, demonstrating that these businesses are well-equipped with the necessary industry data. Technology readiness is another important characteristic across the clusters. In Clusters 1, 2, 3, and 4, MSMEs show Yes for technology readiness, indicating that all groups are prepared to adopt and leverage technology for their operations. This demonstrates the importance of technological adoption for MSMEs across different groupings, suggesting that technology is a priority for enhancing business performance.

The number of businesses in each cluster differs significantly between the unweighted and weighted K-Means analyses. In the weighted analysis, Cluster 1 has 44 members, Cluster 2 has 24 members, Cluster 3 has 10 members, and Cluster 4 has 28 members. This distribution shows how the application of weights impacts the formation of clusters, resulting in more distinct and nuanced groupings compared to the unweighted dataset. The weighted data helps to highlight key differences and identify subgroups that may not be as easily discernible in unweighted analysis.

Cluster profiling provides valuable insights into the unique characteristics of MSMEs by examining various factors such as income, owner education, business experience, awareness, cost management, information access, and technology readiness. The detailed breakdown of these attributes across clusters reveals important patterns that can inform strategic decisions. For example, while all clusters emphasize awareness and cost, certain clusters, like Cluster 2, show a younger demographic with newer businesses, which may require tailored support for growth. The weighted K-Means analysis offers a more refined segmentation, allowing for a deeper understanding of the nuances within MSMEs. This approach helps to identify priority areas such as improving business knowledge or addressing financial challenges based on the specific characteristics of each cluster, thereby aligning business strategies with the unique needs of different MSME groups.

5.4 Development Strategy Based on Halal Certification for MSMEs

This strategy provides a holistic and integrated approach for all MSME clusters, addressing their specific needs and challenges. By implementing the right strategies, MSMEs across all clusters can strengthen their market position, especially in facing halal certification demands and sustainable business development.

CODE	Criteria	Cluster			
		1	2	3	4
Z1	Monthly Income of the MSMEs	500.000 - 5.000.000	5000.000 - 10.000.000	500.000 - 5.000.000	500.000 - 5.000.000
Z2	Last Education of MSMEs Owner	Senior High	Undergraduate	Junior High	Senior High

CODE	Criteria	Cluster			
		1	2	3	4
Z3	Length of Business	5 - 10 Years	1-5 Years	5 - 10 Years	5 - 10 Years
Z4	MSMEs Owner Ages	32 - 39 Years	25 -32 Years	32 - 39 Years	32 - 39 Years
Z5	Awareness	Yes	Yes	Yes	Yes
Z6	Knowledge of getting the certification	No	No	No	Yes
Z7	Cost	Yes	Yes	Yes	Yes
Z8	Information Access to get/obtain the halal certification	Yes	No	Yes	Yes
Z9	Technology Readiness	Yes	Yes	Yes	Yes

Among the various criteria assessed in the clustering analysis of MSMEs' readiness for halal certification, the most prominent and standout factor is:

1. Knowledge of How to Obtain Halal Certification (Z6)

   This is the most critical gap, as three out of the four clusters (Clusters 1, 2, and 3) show a lack of knowledge about the process of obtaining halal certification. Only Cluster 4 demonstrates adequate knowledge. Without a clear understanding, MSME owners may feel overwhelmed by the requirements or avoid pursuing certification altogether.

   Strategy:

   - Implement educational campaigns through BPJPH, Dinas Perindagkop UKM, and local universities.
   - Offer free training sessions and socialization workshops targeted specifically at MSME owners in Clusters 1, 2, and 3.
   - Develop simple guidebooks or infographics to break down the halal certification process.

2. Access to Information (Z8)

   Cluster 2 lacks access to information regarding how to get halal certified, while the rest have good access. This indicates an information gap that may stem from limited exposure to formal support channels, weak digital literacy, or lack of integration into community networks.

   Strategy:

   - Improve digital outreach and social media presence of halal certification information, especially through social media, where MSMEs frequently engage.
   - Collaborate with local community leaders or RT/RW to distribute printed materials in areas with low internet penetration.
   - Integrate information about halal certification into existing MSME support programs (e.g., business assistance or financial literacy).

3. Length of Business (Z3)

   Cluster 2 also includes newer MSMEs (1–5 years old), while the other clusters consist of more experienced businesses (5–10 years). Newer businesses may face more uncertainty, lack strategic planning, and have less administrative or regulatory knowledge.

   Strategy:

   - Provide mentorship or coaching programs by connecting young MSMEs with more experienced businesses from Clusters 1, 3, and 4.
   - Create starter kits for new businesses that include halal certification as part of their early development roadmap.
   - Encourage participation in business incubators or cooperative forums that support halal certification readiness from the ground up.

CHAPTER VI
CONCLUSION AND SUGGESTION

6.1 Conclusion

The following is a closing and conclusion related to this research:

1. The weighted clustering analysis using the K-Means algorithm successfully identified key factors influencing MSMEs in Yogyakarta City in their pursuit of halal certification. By incorporating weights, the analysis revealed that awareness, knowledge, information access, and technology readiness were the most significant factors determining the readiness of MSMEs for obtaining halal certification. The results indicated that MSMEs with higher awareness and technology readiness were better positioned to meet the certification requirements. The weighted clustering analysis further enabled the grouping of MSMEs based on their level of readiness for halal certification, which was determined by various criteria variables. The K-Means algorithm, after weighting, resulted in four distinct clusters:
   - Cluster 1: Mid-Income Established Entrepreneurs – MSMEs with long operational years and good awareness and technology readiness, but lacking knowledge about halal certification.
   - Cluster 2: High-Income Emerging Entrepreneurs – MSMEs with higher income and education, good awareness and technology readiness, but limited access to halal certification information.
   - Cluster 3: Mid-Income Traditionalists – MSMEs with low education levels, decent awareness, but struggling with both knowledge and information access.
   - Cluster 4: Mid-Income Modern Entrepreneurs – The readiest cluster, showing strong performance across all key variables. This cluster only emerged after applying weighted clustering, revealing its previously hidden presence in the dataset.
2. Based on the clustering results, three criteria clearly stand out as potential barriers to halal certification among MSMEs in Yogyakarta: knowledge about obtaining halal certification (Z6), access to information (Z8), and length of business (Z3).

Each of these factors requires tailored strategies to ensure that MSMEs can effectively prepare for and comply with the mandatory halal certification by 2024. The three most critical gaps identified in this research are the lack of knowledge about how to obtain halal certification (Z6), limited access to information (Z8), and shorter business experience (Z3). Knowledge of the certification process is the most significant issue, as Clusters 1, 2, and 3 all show a lack of understanding, which could discourage MSME owners from pursuing halal certification due to confusion or fear of complex requirements. To address this, educational campaigns should be implemented through BPJPH, Dinas Perindagkop UKM, and local universities, alongside free training sessions, workshops, and the development of simple guidebooks or infographics. In terms of information access, Cluster 2 stands out for its deficiency, likely caused by low digital literacy or limited engagement with formal support networks. Strategies should focus on enhancing digital outreach via popular platforms such as WhatsApp, Instagram, and TikTok, working with local leaders to distribute printed materials, and embedding halal certification info into existing MSME support programs. Additionally, Cluster 2's relatively young businesses (1–5 years) face challenges due to lack of experience, strategic direction, and familiarity with regulations. This can be mitigated through mentorship from more established MSMEs, starter kits that include halal certification guidance, and participation in incubator or cooperative forums. These three interrelated problems point to the importance of a layered, inclusive support system that empowers MSMEs at various stages of development to achieve halal certification and remain competitive in Indonesia’s expanding halal economy.

6.2 Suggestion

In summary, the weighted clustering analysis with the K-Means algorithm provided valuable insights into the factors that influence MSMEs in Yogyakarta City’s pursuit of halal certification.

It also successfully identified distinct groupings based on varying levels of readiness, offering a nuanced understanding of MSMEs' needs and challenges in this context. To support MSMEs in achieving halal certification readiness, local government agencies such as BPJPH, Dinas Perindagkop UKM, and academic institutions should collaborate to create a comprehensive Halal Certification Readiness Program. This program should include:

1. Knowledge Empowerment:
   - Develop easy-to-understand guidebooks and visual infographics about the halal certification process.
   - Organize regular offline and online training workshops focused on Clusters 1, 2, and 3.
2. Information Accessibility:
   - Enhance digital communication by using MSME-friendly social media platforms
   - Partner with community leaders to distribute printed materials in areas with low internet access.
3. Support for Young MSMEs:
   - Launch mentorship programs pairing experienced MSMEs (from Clusters 1 and 4) with newer businesses (Cluster 2).
   - Provide halal certification starter kits as part of MSME onboarding in business incubators or cooperatives.

This multi-layered, cluster-based approach ensures that support is targeted, inclusive, and aligned with the actual challenges faced by MSMEs accelerating their halal certification journey and enhancing their competitiveness in the halal economy.

REFERENCES

25-Article Text-75-2-10-20190223. (n.d.).

551575205. (n.d.).

Abdulhafedh, A. (n.d.). *Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation*.

Abdulkadium, A. M., Shekan, R. A. A., & Hussain, H. A. (2022). Application of Data Mining and Knowledge Discovery in Medical Databases. *Webology*, *19*(1), 4912–4924. https://doi.org/10.14704/WEB/V19I1/WEB19329

Adistya, A. P., Lutfiyani, N., Tara, P., Adriyan, R., & Rosyani, P. (2023). *Klasterisasi Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Kelulusan Mata Kuliah Mahasiswa*. *2*(8).

Aggarwal, D., & Sharma, D. (2019). *Application of Clustering for Student Result Analysis*. *7*(6).

Arifah, I. D. C., Prasanna, F. M., & Aziz, M. A. (2022). ANALISIS DAMPAK MONETER, KESIAPAN TEKNOLOGI, DAN ANCAMAN KEAMANAN MATA UANG DIGITAL BANK SENTRAL (CBDC). *Journal of Digital Business and Innovation Management*, *1*(2), 79–100. https://doi.org/10.26740/jdbim.v1i2.50488

Arifin, R., Ningsih, A. A. T., & Putri, A. K. (2021). *THE IMPORTANT ROLE OF MSMEs IN IMPROVING THE ECONOMY*. *24*(6).

Asyari, A. H., & Zagladi, A. N. (2020). Analisis Segmen Pasar Produk Handphone Berbasis Android di Lingkungan Mahasiswa di Banjarmasin. *Jurnal Riset Inspirasi Manajemen dan Kewirausahaan*, *4*(1), 9–15. https://doi.org/10.35130/jrimk.v4i1.73

Benmoussa, K., Laaziri, M., Khoulji, S., Kerkeb, M. L., & Yamami, A. E. (2019). AHP-based Approach for Evaluating Ergonomic Criteria. *Procedia Manufacturing*, *32*, 856–863. https://doi.org/10.1016/j.promfg.2019.02.294

Cui, M. (n.d.). *Introduction to the K-Means Clustering Algorithm Based on the Elbow Method*.

Dalmia, A., & Sia, S. (2021). *Clustering with UMAP: Why and How Connectivity Matters* (No. arXiv:2108.05525). arXiv. https://doi.org/10.48550/arXiv.2108.05525

Document. (n.d.).

Document-1. (n.d.).

Emovon, I., & Oghenenyerovwho, O. S. (2020). Application of MCDM method in material selection for optimal design: A review. *Results in Materials*, *7*, 100115. https://doi.org/10.1016/j.rinma.2020.100115

Gramfort, A., & Michel, V. (n.d.). *296 PUBLICATIONS 107,207 CITATIONS SEE PROFILE*.

Hartama, D., Perdana Windarto, A., & Wanto, A. (2019). The Application of Data Mining in Determining Patterns of Interest of High School Graduates. *Journal of Physics: Conference Series*, *1339*(1), 012042. https://doi.org/10.1088/1742-6596/1339/1/012042

Ihram, I. (2020). PERSEPSI KONSUMEN TERHADAP SAYURAN ORGANIK DAN PENGARUHNYA TERHADAP MINAT BELI DI KOTA MATARAM. *AGROTEKSOS:*

Agronomi Teknologi dan Sosial Ekonomi Pertanian, *29*(2), 70. https://doi.org/10.29303/agroteksos.v29i2.406

Indraswari Octa Nur Audia & Wahyu Dwi Warsitasari. (2023). Pengaruh Pengetahuan, Sikap Dan Pengalaman Keuangan Terhadap Perilaku Manajemen Keuangan Pada Pelaku UMKM Di Kabupaten Tulungagung. *PESHUM : Jurnal Pendidikan, Sosial dan Humaniora*, *2*(3), 385–401. https://doi.org/10.56799/peshum.v2i3.1498

Khan, A. U., & Ali, Y. (2020). ANALYTICAL HIERARCHY PROCESS (AHP) AND ANALYTIC NETWORK PROCESS METHODS AND THEIR APPLICATIONS: A TWENTY YEAR REVIEW FROM 2000-2019: AHP & ANP techniques and their applications: Twenty years review from 2000 to 2019. *International Journal of the Analytic Hierarchy Process*, *12*(3). https://doi.org/10.13033/ijahp.v12i3.822

Kristanto, B., Turmudi Zy, A., & M. Fatchan. (2023). Analisis Penentuan Karyawan Tetap Dengan Algoritma K-Means Dan Davies Bouldin Index. *Bulletin of Information Technology (BIT)*, *4*(1), 112–120. https://doi.org/10.47065/bit.v4i1.521

Markusheski, L., Zdravkoski, I., Andonovski, M., & Jovanoska, A. (n.d.). *KNOWLEDGE DISCOVERY DATABASES (KDD) PROCESS IN DATA MINING*.

Palariya, P. (n.d.). *Project report submitted in partial fulfillment of the requirement for the degree of Bachelor of Technology in*.

Rachman, H. A. (2023). *Akuntansi, Fakultas Ekonomi dan Bisnis, Universitas Brawijaya, Indonesia*. *2*(4).

Sari, M., & Arianto, T. (n.d.). *PENGARUH ROTASI DAN MUTASI KERJA TERHADAP KINERJA KARYAWAN PTPN 7 CABANG BENGKULU*.

Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-Means Clustering Algorithm. *IEEE Access*, *8*, 80716–80727. https://doi.org/10.1109/ACCESS.2020.2988796

Sukmadewanti, I., Arifudin, R., & Sugiharti, E. (2018). Use of K-Means Clustering and Analytical Methods Hierarchy Process in Determining the Type of MSME Financing in Semarang City. *Scientific Journal of Informatics*, *5*(2), 148–158. https://doi.org/10.15294/sji.v5i2.16221

Theresia, W., & Rostiana, R. (2024). Pengaruh literasi keuangan, pengetahuan investasi, dan faktor sosial terhadap minat berinvestasi saham Generasi Z. *Jurnal Manajemen Bisnis dan Kewirausahaan*, *8*(4), 881–894. https://doi.org/10.24912/jmbk.v8i4.31640

Trianda, F. Y. (n.d.). *PENGARUH LABEL HALAL, KESADARAN HALAL DAN HARGA TERHADAP MINAT BERBELANJA ONLINE PRODUK MAKANAN: PERAN KARAKTERISTIK DEMOGRAFI SEBAGAI PEMODERASI (STUDI PADA MAHASISWA FEBI UNIVERSITAS ISLAM NEGERI AR RANIRY)*.

Wang, X., & Xu, Y. (2019). An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index. *IOP Conference Series: Materials Science and Engineering*, *569*(5), 052024. https://doi.org/10.1088/1757-899X/569/5/052024

APPENDICES

A. Questionnaire (Google Form)

Pengelompokan UMKM Berdasarkan Kesiapan Dalam Sertifikasi Halal di Daerah Istimewa Yogyakarta

Assalamualaikum Wr. Wb.

Responden yang Terhormat,

Perkenalkan Kami:
- Antony Sayeed Muhammad (20522060)
- Ade Fikry Mustofa (20522282)
- Fika Rachma Adelia (20522220)
- Husen Fadholi (20522276)
- Naoval Harief (20522284)

Merupakan Mahasiswa semester akhir Jurusan Teknik Industri Program Sarjana, Fakultas Teknologi Industri, Universitas Islam Indonesia. Kami sedang melakukan penelitian Tugas Akhir mengenai "Pengelompokan UMKM Berdasarkan Kesiapan Dalam Sertifikasi Halal di Daerah Istimewa Yogyakarta". Pada Kesempatan ini, kami memohon kerjasama Bapak/Ibu untuk kesediaannya dalam mengisi dan menjawab seluruh pertanyaan yang ada pada penelitian ini.

Untuk Itu diharapkan Para Responden dapat memberikan jawaban yang sebenar-benarnya dalam membantu penelitian ini. Peneliti akan merahasiakan identitas Responden, semua data serta hasil dari kuesioner responden.

Atas Perhatiannya Terimakasih.

Wassalamualaikum Wr. Wb.

* Indicates required question

1. Saya mengisi kuisioner ini dengan sukarela tanpa adanya paksaan *

Mark only one oval.

- Ya
- Tidak

Demografi UMKM

2. Nama UMKM *

3. Kabupaten/kota UMKM *

Mark only one oval.

- Yogyakarta
- Sleman
- Bantul
- Gunung Kidul
- Kulon Progo

4. Lama Usaha *

5. Produk UMKM *

6. Jumlah Produk *

Mark only one oval.

- Tunggal
- Banyak

7. Usia Pemilik UMKM *

Mark only one oval.

- <17 tahun
- 17-25 tahun
- 25-32 tahun
- 32-39 tahun
- 39-46 tahun
- >47 tahun

8. Gender Pemilik UMKM *

Mark only one oval.

- Laki-laki
- Perempuan

9. Pendidikan Terakhir Pemilik UMKM *

Mark only one oval.

- SD
- SMP
- SMA
- Sarjana
- Tidak Sekolah

10. Penghasilan perbulan UMKM *

Mark only one oval.

- <500.000
- 500.000-5.000.000
- 5.000.000-10.000.000
- 10.000.000-15.000.000
- >15.000.000
- >25.000.000
- >50.000.000

***Knowledge* (Pengetahuan)**

11. Pernah mengikuti pelatihan? *

Mark only one oval.

- Ya
- Tidak

12. Mengetahui syarat-syarat pengajuan sertifikasi halal? *

Mark only one oval.

- Ya
- Tidak

13. Mengetahui cara mengajukan sertifikasi halal? *

Mark only one oval.

- Ya
- Tidak

14. Mengetahui dokumen apa yang harus dimiliki untuk mengajukan sertifikasi halal? *

Mark only one oval.

- Ya
- Tidak

15. Mengetahui kesulitan dalam mengajukan sertifikasi halal dan cara mengatasinya? *

Mark only one oval.

- Ya
- Tidak

***Awareness* (Kesadaran)**

16. Kesadaran akan manfaat yang didapat dari sertifikasi halal? *

Mark only one oval.

- Ya
- Tidak

17. Kesadaran akan produk makanan dan minuman harus bersertifikat halal? *

Mark only one oval.

- Ya
- Tidak

18. Apakah anda menjamin produk yang belum memiliki sertifikasi halal itu sudah aman? *

Mark only one oval.

- Ya
- Tidak

19. Kesadaran pentingnya sertifikasi halal bagi masyarakat? *

Mark only one oval.

- Ya
- Tidak

20. Apakah produk yang Anda jual harus memiliki sertifikasi halal? *

Mark only one oval.

- Ya
- Tidak

Sikap

21. Apakah makanan dan minuman halal itu penting? *

Mark only one oval.

- Ya
- Tidak

22. Apakah memiliki sertifikat halal itu penting? *

Mark only one oval.

- Ya
- Tidak

23. Apakah merasa aman jika menawarkan makanan dan minuman halal? *

Mark only one oval.

- Ya
- Tidak

24. Apakah memprioritaskan makanan dan minuman halal? *

Mark only one oval.

- Ya
- Tidak

25. Apakah menyediakan makanan dan minuman halal adalah hal positif? *

Mark only one oval.

- Ya
- Tidak

***Technology Performance* (Kesiapan Teknologi)**

26. Apakah pemilik atau staff memiliki kemampuan menggunakan komputer? *

Mark only one oval.

- Ya
- Tidak

27. Apakah pemilik atau staff memiliki kemampuan mengakses internet? *

Mark only one oval.

- Ya
- Tidak

28. Apakah ada orang di sekitar UMKM yang dapat membantu mengoperasikan komputer dan mengakses internet? *

Mark only one oval.

- Ya
- Tidak

29. Apakah UMKM yang dikelola memiliki komputer pribadi? *

Mark only one oval.

- Ya
- Tidak

30. Apakah UMKM yang dikelola memiliki akses internet pribadi? *

Mark only one oval.

- Ya
- Tidak

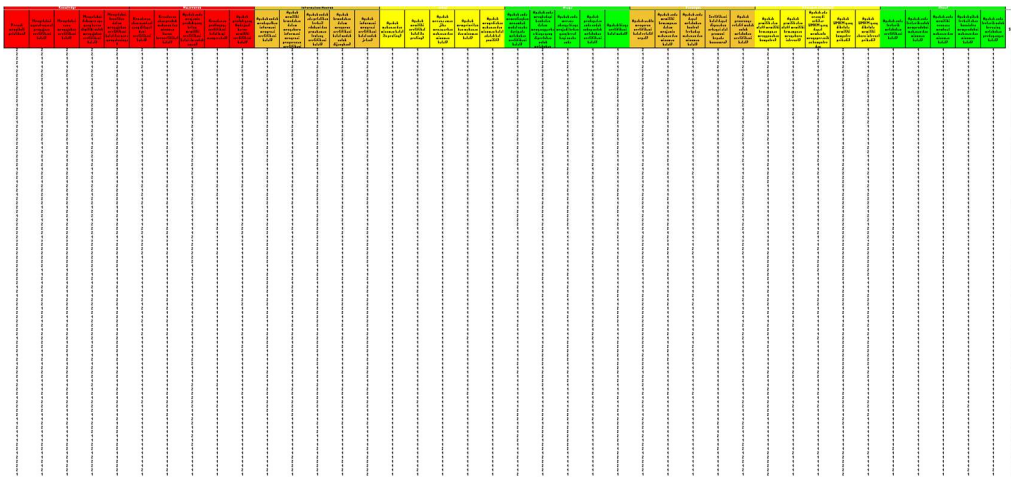
This content is neither created nor endorsed by Google.

Google Forms

B. Clustering MSMEs Data Recapitulation

	Demografi				Pengetahuan				
Nomor	Nama UMKM	Lama Usaha	Pendidikan Terakhir Pemilik UMKM	Penghasilan perbulan UMKM	Pernah mengikuti pelatihan?	Mengetahui syarat-syarat pengajuan sertifikasi halal?	Mengetahui cara mengajukan sertifikasi halal?	Mengetahui dokumen apa yang harus dimiliki untuk mengajukan sertifikasi halal?	Mengetahui kesulitan dalam mengajukan sertifikasi halal dan cara mengatasinya?
1	Dago Coffee	2	4	5	0	0	0	0	0
2	pecel	2	4	1	0	1	1	1	1
3	Es Kepal Milo Nikmat Rasa	2	4	2	0	1	1	1	1
4	7Rasa Catering	4	4	5	1	1	1	1	0
5	Funny Warmindo	2	1	2	0	0	0	0	1
6	sate kambing pak udin	3	3	2	0	0	0	0	0
7	WestPush Coffee	2	4	2	0	0	1	1	1
8	Mie Ayam Afni	3	4	5	0	1	1	0	0
9	Dawet Jeruz Yudhistira	1	4	2	0	1	1	1	1
11	Nasi Berkat	1	4	2	0	1	1	1	1
12	Pawon Adyasta	2	3	1	0	0	0	0	0
13	Dapur Erlin	2	3	2	1	1	1	0	0
14	Bubur Kacang Ijo	6	1	1	0	0	0	0	0
15	Siomay Ken Ken	2	2	2	0	0	0	0	0
16	Pecel lele lamongan	2	2	4	0	0	0	0	0
17	Rizky kebab dan burger	2	2	2	1	1	1	1	1
18	gepuek bao	2	4	2	0	1	1	1	1
19	Creamy Fruity Jelly	1	4	1	0	0	0	0	1
20	Pinjuit Pedas	1	3	2	0	0	0	0	0
21	Ceker tanpa tulang	1	3	2	0	0	0	1	0
22	es teh satu rasa	1	4	4	0	1	1	1	0
23	3ya Block	2	4	2	0	1	0	0	0
24	mie fil fil	1	4	2	0	1	1	0	0
25	bana gimbab	3	3	3	0	0	0	0	0
26	Jumpia rebung ulala	2	4	2	0	1	1	1	0
27	catering intan & fika	4	3	2	0	0	0	0	0
28	Angkringan Mbah Nono	4	4	2	1	1	1	1	1
29	nisol mayo	1	3	1	0	0	0	1	0
30	Angkringan	7	2	2	0	0	0	0	1
31	Mie Ayam Om Totok	7	3	2	1	1	1	1	1
32	mie ayam tim tam	3	3	2	0	0	0	0	0
33	Madhep Mantep	5	2	5	0	0	0	0	0
34	Ayam goreng bu hartin	3	4	3	1	0	1	1	0
35	Warteg Pempek Cik Ana	2	4	2	0	0	0	0	0
36	Mie Ayam & Bakso Karomah	4	4	2	0	0	0	0	0
37	Pecel Lele Remaja 77	2	3	2	0	1	1	1	1
38	donat idola	2	4	2	0	0	0	0	1
39	Sego Bakar 88	2	3	2	1	0	0	0	0
40	Go Angkringan	2	4	2	0	1	1	1	1
41	Juice & Sup Buah Antartika	3	3	2	0	0	0	0	0
42	ayam geprek bu hasong	5	2	2	0	0	0	0	0
43	ayam geprek bu made	5	2	3	0	0	0	0	0
44	ayam geprek bu nan	6	1	3	0	1	0	1	1
45	Warmindo Kanmia	4	3	2	0	0	0	0	1
46	bubur ayam kabita	3	3	2	0	0	0	0	0
47	sate kambing mbah nanny riyanto	2	3	2	0	0	0	0	0
48	warung soto bu tiar	2	2	2	0	0	0	0	0
49	bubur susssum bu sri suparmi	4	3	2	0	0	0	0	1
50	Pempek kenangan	3	2	3	0	0	0	0	0
51	My crepezzz	1	2	1	0	0	0	0	0
52	bakmi jawa handayani	4	2	3	0	1	1	0	1
53	baureng jadul	2	1	2	0	1	0	1	1
54	Warteg Bu Sum	4	3	2	0	0	0	0	1
55	Kerinik nisanur as-syifa	2	3	2	0	1	1	1	1
56	Muda Berkah Creative	2	3	2	1	0	0	0	0
57	Sate madura H Afifi	4	3	3	0	0	1	1	1
58	gadog yo nanie	3	2	2	1	1	1	1	1
59	es dawet dusun ndati	2	2	2	0	0	0	0	0
60	siomay barokah kota wak	3	2	2	0	0	0	0	0
61	Colonial Rice Bowl	3	3	3	0	1	0	0	1
62	Bubur Kacang Gerobak Biru	2	1	1	0	0	0	0	0
63	ebkicopa	3	4	3	0	0	1	1	0
64	Dyok Sola	1	4	1	0	1	1	0	0
65	Bakpia 2S	7	2	2	0	0	0	0	0
66	Lontong opor busiam	4	2	4	0	1	1	0	1
67	es oyen balagor bandung magewo	3	3	2	0	0	0	0	0
68	Jcbyuuoo	2	3	2	1	0	0	0	0
69	bubur ayam khas jakarta (palagan)	2	4	2	0	0	1	1	0
70	ikan pepes mbah im	3	1	2	0	0	0	0	0
71	es dogger balai yasa	4	4	3	0	1	0	1	0
72	warung nolok mbak yuni kota batu	5	3	4	1	1	1	1	1
73	kereta kopi kota batu	2	4	3	0	0	0	0	0
74	Ilty yuliastuti	3	3	2	1	1	1	1	1
75	Sate Kamnia	4	2	2	0	0	0	0	1

Nomor	Demografi				Pengetahuan				
	Nama UMKM	Lama Usaha	Pendidikan Terakhir Pemilik UMKM	Penghasilan perbulan UMKM	Pernah mengikuti pelatihan?	Mengetahui syarat-syarat pengajuan sertifikasi halal?	Mengetahui cara mengajukan sertifikasi halal?	Mengetahui dokumen apa yang harus dimiliki untuk mengajukan sertifikasi halal?	Mengetahui kesulitan dalam mengajukan sertifikasi halal dan cara mengatasinya?
76	Soto betawi	2	3	2	0	0	0	0	0
77	Marago Mulyo	3	1	2	0	0	0	0	0
78	Ariesgar	2	3	3	1	1	1	1	1
79	Ayamgepeek bu Ana Lodadi	4	3	3	0	1	1	0	1
80	Bakso dan Mie ayam pak satu	5	2	5	0	1	0	1	0
81	Nasi rames bu muji	5	2	2	0	0	0	0	0
82	Spectrum	4	1	3	1	1	1	1	1
83	Bakwan Kawi Deredan	2	1	1	0	0	0	0	0
84	Bakso cuanki dejolan	2	1	1	0	0	0	0	0
85	Gudeg basah bu ani	4	3	3	0	0	1	0	1
86	Bubur kacang ijo pak tit	3	2	3	0	1	1	1	1
87	nayonkitchen.id	2	3	2	1	1	1	1	0
88	Kuliner Tenis Mranggi	2	3	2	1	0	0	0	1
89	Kedai Seminah Kraton	2	3	2	0	0	1	0	0
90	Soto DaginaPa Inn	3	3	2	1	1	1	0	0
91	Francis Pizza	4	4	3	1	1	1	1	1
92	Susu 4	3	4	5	1	1	1	1	1
93	Soto ayam & Sate Kerang Bu Sti	2	4	3	0	0	0	0	0
94	Jan Bajog	2	3	5	0	0	0	0	0
95	Dawet Ayu	7	5	2	1	1	1	1	1
96	Mie Ayam Bakso Gapura	2	2	2	0	1	0	0	1
97	Martabak	2	3	2	0	0	1	0	0
98	Mie Ayam	2	5	2	0	1	0	0	0
99	bakwan kawi fkri	1	3	2	0	0	0	0	0
100	thiwul dan aneka ba tiqui	4	3	2	1	1	1	1	1
101	gado-gado lotek kupat tahu bu gend	2	3	3	1	1	1	1	0
102	cilok barokah	2	3	2	0	1	0	0	0
103	Bakso tusuk payaman	1	3	2	0	0	0	0	0
104	Martabak RR	2	3	2	0	1	1	0	0
105	batagor dengan KPRI	2	2	2	0	1	0	0	0
106	kue pukis	2	1	2	0	0	0	0	0



C. AHP (Pairwise Comparison) Data

E1									
Criteria	Cost	Awareness	Knowledge	Information Access	Technology Readiness	MSMEs Owner Ages	Last Education of MSMEs Owner	Length of Business	Monthly Income of the MSMEs
Cost	1	0.33333333	6	2	3	6	5	6	4
Awareness	3	1	3	2	4	7	6	7	5
Knowledge	0.16666667	0.33333333	1	2	2	4	3	5	2
Information Access	0.5	0.5	0.5	1	2	5	3	6	3
Technology Readiness	0.33333333	0.25	0.5	0.5	1	3	2	4	2
MSMEs Owner Ages	0.16666667	0.14285714	0.25	0.2	0.33333333	1	0.5	2	0.33333333
Last Education of MSMEs Owner	0.2	0.16666667	0.33333333	0.33333333	0.5	2	1	2	0.5
Length of Business	0.16666667	0.14285714	0.2	0.16666667	0.25	0.5	0.5	1	0.25
Monthly Income of the MSMEs	0.25	0.2	0.5	0.33333333	0.5	3	2	4	1
Total	5.78333333	3.06904762	12.2833333	8.53333333	13.5833333	31.5	23	37	18.0833333

E2									
Criteria	Cost	Awareness	Knowledge	Information Access	Technology Readiness	MSMEs Owner Ages	Last Education of MSMEs Owner	Length of Business	Monthly Income of the MSMEs
Cost	1	0.5	0.5	0.5	2	1.5	2	2	1.5
Awareness	2	1	2	4	3	5	4	5	5
Knowledge	2	0.5	1	3	3	4	4	5	6
Information Access	2	0.25	0.33	1	2	3	3	4	3
Technology Readiness	0.5	0.33	0.33	0.5	1	2	3	2	2
Usia Pemilik UMKM	0.67	0.2	0.25	0.33	0.5	1	0.5	1	0.25
Pendidikan Terakhir Pemilik	0.5	0.25	0.25	0.33	0.33	2	1	2	0.5

Lama Usaha UMKM	0.5	0.25	0.25	0.25	0.33	2	0.5	1	0.5
Penghasilan UMKM	0.67	0.2	0.17	0.33	0.5	3	2	2	1
Total	9.84	3.48	5.08	10.24	12.66	23.5	20	24	19.75

E3									
Criteria	Cost	Awareness	Knowledge	Information Access	Technology Readiness	MSMEs Owner Ages	Last Education of MSMEs Owner	Length of Business	Monthly Income of the MSMEs
Cost	1	0.333	0.25	0.5	2	1.5	2	2	1.5
Awareness	3	1	2	4	3	5	4	5	6
Knowledge	4	0.5	1	3	3	4	4	5	7
Information Access	2	0.25	0.33	1	2	3	3	4	3
Technology Readiness	0.5	0.33	0.33	0.5	1	2	3	2	2
Usia Pemilik UMKM	0.67	0.2	0.25	0.33	0.5	1	0.5	1	0.25
Pendidikan Terakhir Pemilik	0.5	0.25	0.25	0.33	0.33	2	1	2	0.5
Lama Usaha UMKM	0.5	0.33	0.25	0.25	0.33	2	0.5	1	0.5
Penghasilan UMKM	0.67	0.17	0.14	0.33	0.5	3	2	2	1
Total	12.84	3.363	4.8	10.24	12.66	23.5	20	24	21.75

E4									
Criteria	Cost	Awareness	Knowledge	Information Access	Technology Readiness	MSMEs Owner Ages	Last Education of MSMEs Owner	Length of Business	Monthly Income of the MSMEs
Cost	1	0.333	0.333	0.5	3	2	3	3	2
Awareness	3	1	2	5	4	7	6	4	7
Knowledge	3	0.5	1	3	4	5	5	5	8
Information Access	2	0.2	0.333	1	2	3	3	4	3
Technology Readiness	0.333	0.25	0.25	0.5	1	2	4	3	2
Usia Pemilik UMKM	0.5	0.142	0.2	0.333	0.5	1	0.333	0.5	0.25
Pendidikan Terakhir Pemilik	0.333	0.167	0.2	0.333	0.25	3	1	2	0.5
Lama Usaha UMKM	0.333	0.25	0.2	0.25	0.333	2	0.5	1	0.5
Penghasilan UMKM	0.5	0.143	0.125	0.333	0.5	4	2	2	1
Total	10.999	2.985	4.641	11.249	15.583	29	24.833	24.5	24.25

E5									
Criteria	Cost	Awareness	Knowledge	Information Access	Technology Readiness	MSMEs Owner Ages	Last Education of MSMEs Owner	Length of Business	Monthly Income of the MSMEs
Cost	1	1	1	1	1	0.5	1	0.2	1
Awareness	1	1	1	1	1	0.5	1	0.5	1
Knowledge	1	1	1	1	1	0.5	1	0.5	0.5
Information Access	1	1	1	1	1	0.5	1	0.5	0.5
Technology Readiness	1	1	1	1	1	1	1	0.5	1
Usia Pemilik UMKM	2	2	2	2	1	1	3	3	1
Pendidikan Terakhir Pemilik	1	1	1	1	1	0.333333 33	1	0.2	1
Lama Usaha UMKM	5	2	2	2	2	0.333333 33	5	1	1
Penghasilan UMKM	1	1	2	2	1	1	1	1	1
Total	14	11	12	12	10	5.666666 67	15	7.4	8

D. Gap statistic Python Coding Script

```
#Weighted K-Means Clustering

import pandas as pd
from sklearn.cluster import KMeans
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import files
from sklearn.metrics import pairwise_distances
# Install umap-learn if it's not already installed
!pip install umap-learn
# Import umap
import umap.umap_ as umap # Changed the import to umap.umap_
# or import umap
# Step 1: Upload Excel file
uploaded = files.upload()  # Upload your Excel file

# Step 2: Load data from the Excel file
file_name = 'Nobobot2.xlsx'  # Replace with the name of your uploaded file
data = pd.read_excel(file_name)

# Step 3: Select columns to be used
features = data[['Z1', 'Z2', 'Z3', 'Z4', 'Z5', 'Z6', 'Z7', 'Z8', 'Z9']]

# Function to calculate Gap Statistic .....
def calculate_gap_statistic(data, k_range, n_refs=10):
    """
    Function to calculate Gap Statistic.
    - data: data to be clustered
    - k_range: range of cluster numbers to test
    - n_refs: number of random reference data for Gap Statistic calculation
```

```
    """
    gaps = []
    for k in k_range:
        # Calculate Wk for the original data
        kmeans = KMeans(n_clusters=k, random_state=42)
        kmeans.fit(data)
        Wk = np.sum(np.min(pairwise_distances(data, kmeans.cluster_centers_,
metric='euclidean'), axis=1))

        # Calculate Wk for random reference data
        Wk_refs = []
        for _ in range(n_refs):
            random_data = np.random.random_sample(size=data.shape)
            kmeans.fit(random_data)
            Wk_random = np.sum(np.min(pairwise_distances(random_data,
kmeans.cluster_centers_, metric='euclidean'), axis=1))
            Wk_refs.append(Wk_random)

        # Gap Statistic
        gap = np.log(np.mean(Wk_refs)) - np.log(Wk)
        gaps.append(gap)

    return gaps

# Step 4: Define the range for the number of clusters
k_range = range(2, 5)  # Adjust this range as needed

# Calculate Gap Statistic for different cluster numbers
gap_values = calculate_gap_statistic(features, k_range)

# Step 5: Visualize Gap Statistic
plt.plot(k_range, gap_values, marker='o')
plt.title('Gap Statistic for Different Numbers of Clusters')
plt.xlabel('Number of Clusters')
```

```
plt.ylabel('Gap Statistic')
plt.xticks(k_range)
plt.grid()
plt.show()

# Determine the optimal number of clusters
optimal_clusters = k_range[np.argmax(gap_values)]
print(f"Optimal Number of Clusters (based on Gap Statistic): {optimal_clusters}")

# Step 6: Perform K-Means Clustering with the optimal number of clusters
kmeans = KMeans(n_clusters=optimal_clusters, random_state=42)
kmeans.fit(features)

# Get cluster labels and centroids
labels = kmeans.labels_
centroids = kmeans.cluster_centers_

# Create a DataFrame for centroids
centroids_df = pd.DataFrame(centroids, columns=features.columns)

# Step 7: Add cluster labels to the original data
data['Cluster'] = labels

# Set options to print the entire dataset
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)

# Display the dataset with cluster labels
print("Final Cluster Assignments:")
print(data)
# Display centroids
print("\nCentroids of the clusters:")
print(centroids_df)
```

```
# Step 8: Visualization using UMAP
# Reduce dimensions to 2 using UMAP
reducer = umap.UMAP(n_components=2, random_state=42)
features_umap = reducer.fit_transform(features)

# Create a DataFrame for UMAP results
umap_df = pd.DataFrame(data=features_umap, columns=['UMAP Component 1', 'UMAP Component 2'])
umap_df['Cluster'] = labels

# Add centroids to UMAP
centroids_umap = reducer.transform(centroids)  # Transform centroids to UMAP space
centroids_umap_df = pd.DataFrame(data=centroids_umap, columns=['UMAP Component 1', 'UMAP Component 2'])
centroids_umap_df['Cluster'] = range(optimal_clusters)  # Assign cluster labels to centroids

# Visualize UMAP results with centroids
plt.figure(figsize=(10, 6))
sns.scatterplot(
    x='UMAP Component 1', y='UMAP Component 2',
    hue='Cluster', data=umap_df,
    palette='Set2', alpha=0.6, s=100
)
sns.scatterplot(
    x='UMAP Component 1', y='UMAP Component 2',
    data=centroids_umap_df,
    color='red', marker='X', s=200, label='Centroids'
)
plt.title('UMAP Visualization of Clusters with Centroids')
plt.xlabel('UMAP Component 1')
plt.ylabel('UMAP Component 2')
plt.legend()
plt.grid()
plt.show()
```