JURNAL ILMIAH TEKNIK INDUSTRI
ISSN: 1412-6869 (Print), ISSN: 2460-4038 (Online)
Journal homepage: http://journals.ums.ac.id/index.php/jiti/index
doi: 10.23917/jiti.v20i2.15374

Classification of Small and Medium Industry in Malang Regency Using K-Means based Clustering Method

Sunday Noya[1a♦], Novenda Kartika Putrianto[1b], Weshley Valentino Salendu[1]

Abstract. *Information in the form of a structured database becomes an inevitable requirement in the management and development of an organization. In Malang Regency, there are thousands of Small and Medium Industries (SMIs) that have not been completely recorded. This makes it difficult for stakeholders to find out the profile and potential of SMIs, especially SMIs that require special attention. This study aims to classify SMIs in Malang Regency by utilizing the K-means clustering method, using 5 clustering variables: the type of business, investment value, business age, product variations, and average monthly income. The data used in this study is primary data which consist of the five variables of the 183 SMIs sample located in Malang Regency. The result is that the SMIs are classified into 3 clusters, and patterned based on investment value, business age, and average monthly income. Nearly 90% of SMIs are grouped into clusters with the lowest specifications which provide an overview of the current condition of SMIs in Malang Regency.*

Keywords: *small and medium industry, K-means, clustering..*

I. Introduction

Information is valuable for every organization in order to analyze performance, evaluate current position and make plans for further development. It is uncontested that nowadays data has become a significant part of life. Therefore, a complete and structured database is a very important requirement. Database is a collection of systematized information that can be easily accessed, managed, utilized and updated. The need for a structured database also includes information on Small and Medium Enterprises (SMEs). Based on data from the Ministry of Cooperatives and Small and Medium Enterprises of the Republic of Indonesia, as of 2018, the number of Medium Small and Micro Enterprises (MSMEs) in Indonesia reached 63 million businesses, 99% of all businesses in Indonesia (depkop.go.id, 2021). With a very large number and with the potential and business characteristics that are disparate from one MSME to another, it is necessary to form a well-classified database.

Small and Medium Industrie (SMIs) is part of Small and Medium Enterprises (SMEs). Industry is defined as an economic activity that converts raw materials into finished or semi-finished goods to increase added value; while the wider scope of SMEs includes marketing, distribution, and other activities such as packaging, labeling, etc. in their core business. These two terms are often used interchangeably. In this article the term SMI is used to emphasize that the business units in this study are business units that have production activities. Differentiating them from SME, which includes business units that only carry out repacking and reselling without any production activities.

In Malang Regency, based on observation and unofficial information, there are thousands of Small and Medium Industries (SMIs) spread throughout the very large area of the Regency. Currently, the Malang Regency Department of Industry and Trade all of the SMIs is still in the process of collecting the complete data of SMIs in their region. With a very large number and wide potential differences and characteristics, a systematic and structured database and data classification system is needed, which makes it easier for stakeholders to access, manage, use, and update the SMI data effectively. Complete,

[1] Department of Industrial Engineering, Faculty of Engineering, Universitas Ma Chung, Jalan Villa Puncak Tidar N-01, Malang, Jawa Timur 65151

[a] email: sunday.alexander@machung.ac.id
[b] email: novenda@msn.com
♦ corresponding author

Submited: 04-08-2021 Revised: 12-12-2021
Accepted: 22-12-2021

systematic and well-classified IKM data can be very useful to provide information to stakeholders regarding the profile and potential of SMIs. This kind of information can then be used for SMIs development strategic planning in Malang Regency or as a basis for other policy making. Sunaryanto in Supriyanto et. al. (2017) claims that the cluster-based SMIs development strategy that utilizes information technology can improve the efficiency and performance of these SMIs. Meanwhile, Akhmad and Susantiaji (2020) classify SMEs data as the basis for making policies for developing the SME sector in their region. Taneo et. al. (2019) made an SMIs innovation model in developing their sustainable competitiveness, states that the SMI community (paguyuban) is one of the promoters of SMI innovation. The initiation of the formation of these communities will be helped if there is data on the classification of SMIs. SMIs with similar characteristics can form communities that support each other in innovation.

With a very large number and very diverse characteristics, classification is one of the best ways to identify needs, organize, assist, and develop SMIs. In Indonesia, several studies have been conducted to map SMI based on various aspects and objectives. Among them are Madyaratry et. al. (2020) who classified SMI banana chips in Lampung based on the sustainability index, Ahmad and Susantiaji (2020) mapped SMEs in Tegal by number per region, Kresnanto (2020) classified SMEs in Bantul based on their characteristics and spatial data, Raharjo (2017), Indriani and Kartini (2018) classified SMEs in Banjar based on the profile and characteristics of SME entrepreneurs, while Supriyanto (2017) classified SMEs in Semarang based on the location and region of SMEs. These studies utilized different classification methods and instruments based on the needs and characteristics of the data.

This study aims to classify SMIs in Malang district by utilizing the K-means clustering method. K-means clustering is the most frequently used unsupervised clustering technique for classifying a given dataset into clusters. It is an iterative optimization method based on the initial partition of objects into clusters (2020). Cluster analysis utilizes an assessment of similarity or dissimilarity for setting points in space to a cluster (2015). There are five clustering variables used in this study. The variables are type of business, investment value, business age, product variety, and average monthly income. This classification is expected to be used as the basis for making SMIs database in Malang Regency. This database can then be used by the Department of Industry and Trade of Malang Regency and other stakeholders as the basis for formulating policies in developing SMIs in Malang Regency and forming communities that support each other in innovation.

II. Research Method

The data used in this study is primary data of business profile, type of business, investment value, business age, product variations and average monthly income of SMIs. The sample in this study was 183 SMIs spread throughout Malang Regency. To find out whether the number of samples is sufficient to represent the population, an effect size calculation was carried out.

Data processing in this study consists of steps as follows:

Stage 1

Calculating effect size. Effect size is the standardized difference between statistical values and parameter values (Cohen, J. 1988). The smaller the effect size, the smaller the difference between statistical value and the parameter value. Calculating the effect size can be used if the researcher is not able to know the value of a parameter. According to Cohen (1988), there are three effect size categories for the independent average test, that is small effects with effect size values = 0.2, medium effects with effect size values = 0.5, and large effects with effect size values = 0.8. In this study, the effect size was calculated using g-power software with parameters such as significance level (α) = 0.05; sample size (n) = 183; and Power (1-β). Cohen (1988) explained that research should be designed in such way that it has at least 80%

power to detect an effect. So this study uses a power of 0.8.

Stage 2

Data cleaning. Data cleaning is done to maintain the quality of the obtained data. The data should be cleaned and verified from outlier values that different significantly from average data and disparity of attributes which can reduce the efficiency of data mining [14]. Data whose attributes deviated by ±3σ from the mean [15] are deemed to be abnormal and should be removed from the sample. At this stage, cleaning is carried out when the data has been processed using Microsoft Excel.

Data normalization. The data set's variables have distinct characteristics and fluctuations, which might lead to the computation of bias distance. Because of the changes, utilizing the original data in clustering analysis may impact the clustering outcome. As a consequence, the original data should be transformed to a synthetic data set that represents the original data's normalization result. The conversion procedure for normalization of data point is as follows:

$$d_{ij} = \frac{D_{ij}-D_j^{min}}{D_j^{max}-D_j^{min}} \quad \text{...} (1)$$

Where,

D_{ij} is the original value of ith data point in jth dimension,

D_j^{min} is the minimum value of all data points in jth dimension,

D_j^{max} is the maximum value of all data points in jth dimension. i ε {1,2,3,...,S}, { }, j ε {1,2,3,...,M} and

S is the number of data points and M is number of variables. The transformed data will become continuous number within [0,1] range.

K-Means based clustering. Clustering is a method for finding and grouping data that have similar characteristics. The clustering process in this study uses the K-Means method which is generally carried out with the following algorithm:

1. Determining the number of clusters k. The number of cluster used is 3.
2. Initiating a randomly assigned initial cluster center in each cluster.
3. Calculating the distance of each input data xi to each centroid μi using the Euclidean distance formula to find the closest distance from each data to the centroid. Euclidean distance equation:

$$d(x_i, \mu_i) = \sqrt{(x_i - \mu_i)^2} \quad (2)$$

4. Classifying each data based on proximity to the centroid (the smallest distance).
5. Updating the centroid value. The new centroid value is obtained from the average of the cluster in question using the formula:

$$C_k = \frac{1}{n_k} \sum d_i \quad (3)$$

nk is the amount of data in the cluster and di is the sum of the distance values included in each cluster.

6. Repeating steps 2 to 5 until the members of each cluster do not change anymore.

III. Result and Discussion

Stage 1: Calculation of effect size number using g-power software with parameters = 0.05; power (1-β) = 0.8 and n = 183, obtained 0.208. The effect size of 0.208 is categorized as a small effect. Thus, it can be concluded that the statistical or sample mean value has a small effect or is not much different from the parameter or population mean value. In other words, the number of samples collected can represent the population parameters.

Stage 2: Of the five variables used for clustering, it was found that the data is patterned based on 3 variables, investment value, business age, and average monthly income. From the 183 SMIs that were sampled, 3 clusters were set, the green cluster, the yellow cluster and the red cluster. Those grouped in the green cluster are 13 SMIs with an investment value of Rp. 100,000,000 – Rp. 500,000,000, with business age of more than 20 years, and an average monthly income of Rp 50,000,000 – Rp. 70,000,000. The second cluster is the yellow cluster consisting of 7 SMIs with investment values ranging from Rp. 50,000,000 – 80,000,000, with an operating period of 10 – 18 years, and an average monthly income of Rp. 15,000,000 – Rp. 25,000,000. The last cluster is the red cluster with the highest number of members,163 SMIs which has the smallest investment value, which is Rp. 1.000.000 – Rp. 35,000,000, the business age is under 10 years, and the monthly income is only between Rp. 300,000 – Rp. 12,000,000. Almost 90% of SMIs are grouped into the red cluster.

The green cluster is the SMIs group that has the highest average investment value and the largest average monthly income compared to other clusters. The results of the clustering also show that all of the SMIs in the green cluster has been able to survive for more than 20 years. In terms of business performance, it can be concluded that the SMIs in the green cluster experienced quite good business development. The longer the business operates, the more experience it has, the better the ability to detect, manage and maintain the market. This also affects operating income. With more than 20 years of business experience, SMIs in the green cluster have a greater competitive advantage and higher business sustainability.

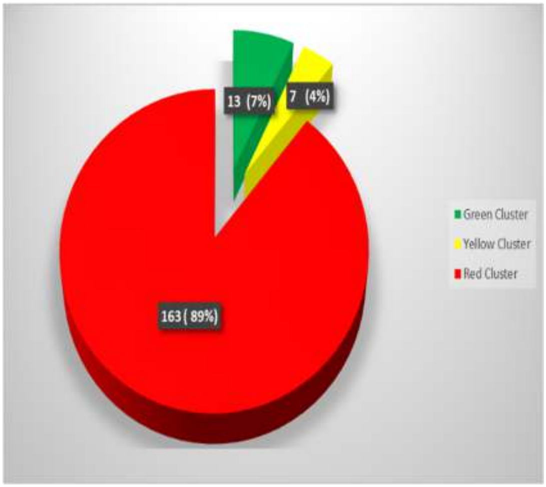

Figure 1. Clustered SMIs Percentage

The yellow cluster consists of SMIs with medium investment and income values and have been running for 10-18 years. The SMIs in the yellow cluster have experienced quite good business development, it can be assumed from their high average monthly income. Although less experienced than SMIs in the green cluster, SMIs in the yellow cluster have relatively had enough time to adapt to various business conditions. The ability to manage the market in this group is also assumed to be very good. The sustainability of SMIs in this cluster can be seen from their ability to survive in a relatively long period of time.

The red cluster consists of SMIs with the lowest investment value and average monthly income. Almost all SMIs in Malang Regency are grouped in this cluster. Because they still have a business age of under 10 years, even quite a lot of them are still 1 or 2 years old, the business stability of SMIs in this group cannot be predicted. The variation between SMIs in this group is very large, but it is clear that with the limited amount of investment and relatively short business life, it can be assumed that SMIs in this group do not have good business experience and skill. This situation, of course, has an impact on the ability to manage the market and the amount of income each month. The business performance of SMIs in this group is still very low and its sustainability has not been tested.

The majority of SMIs in Malang Regency are still in a very weak condition, it can be said that this situation represents the general condition of SMIs in Indonesia. In contrast to the first two groups, which have been able to develop themselves independently, the SMIs in the last cluster are clearly still in need of external support and intervention. The external parties whose involvement is needed are the government, academia, and the SMI community [12]. These supports can be in various forms such as business incentives, soft loans, business facilities, ease of licensing, all of which can be provided by the central and regional governments. Academics can contribute through training, mentoring, and research on business development strategies. No less important is the support from the SMIs community which can be provided in the form of peer assistance, experience sharing, and a support system.

From the distribution of observed data in the three clusters, consistent patterns of comparison between investment value, business age, and average monthly income were found. The patterns are then confirmed using the scatter diagrams as seen in Figure 2.

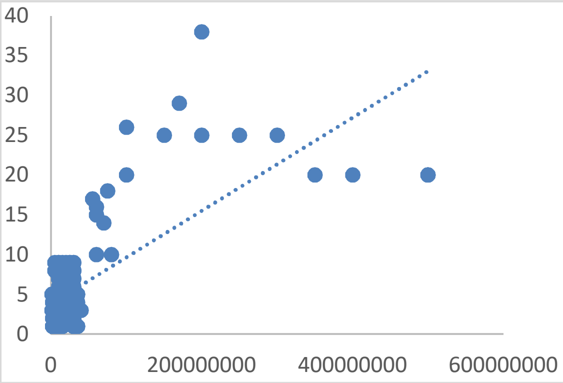

Figure 2. Scatter Diagram: Investment Value vs. Business Age

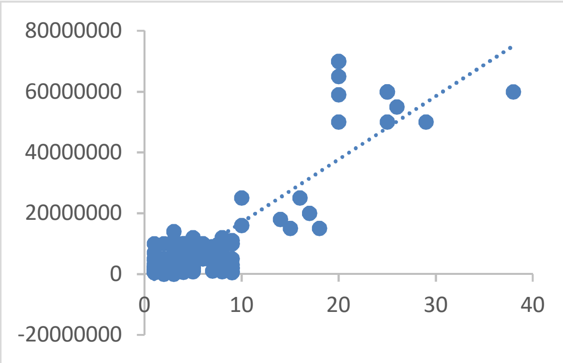

Figure 3. Scatter Diagram: Business Age vs. Avarage Montly Income

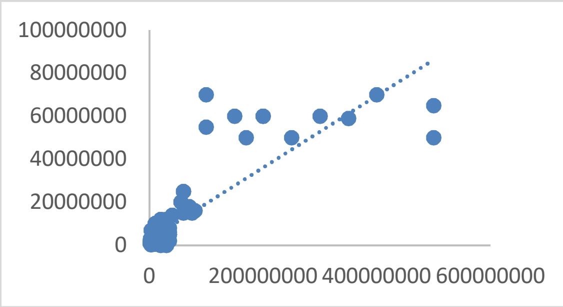

Figure 4. Scatter Diagram: Investment Value vs. Avarage Monthly Income

From the three scatter diagrams above, it can be seen that investment value and business age are directly proportional to average monthly income. This means that the greater the investment value, the longer the business will sustain. The greater the investment value and business age of an SMI, the greater the monthly income. The pattern formed shows the logical consequences of the existence of an SMI. The longer the business age, the business actors have sufficient business experience and knowledge to manage businesses, formulate business strategies, create business innovations, and identify, manage, and expand markets. This condition has a positive impact on business performance which can be seen from operating income [13]. On the other hand, the SMI group with a small investment value also has little business experience, and this results in relatively poor business performance.

The profile and pattern formed from this clustering can provide a clear picture for stakeholders, especially the government regarding the conditions and specifications of SMIs in their region. Aid projects, development projects, and mentoring projects should be focused on SMIs in the lowest cluster. The majority of SMIs with low investment value, minimal business experience, small business income, and poor business performance really need the government's attention.

IV. Conclusion

We have segmented SMIs by using k-means based clustering into 3 parts . SMIs in the three clusters are patterned based on investment value, business age and average monthly income. The pattern formed shows the logical consequences of the existence of an SMI. The bigger the investment value and business age of an SMI, the bigger the monthly income income. The first cluster consists of 7% SMIs with large investment value, business age, and average monthly income. This group has experience and knowledge of good business management, good sustainability, and good business performance. The second cluster consists of 4% SMIs with medium investment value, business age, and average monthly income. The SMIs group in this cluster has relatively good sustainability and business performance. The third cluster, which is the largest cluster with 89% of all SMI members, consists of SMIs with moderate investment value, business age and average monthly income. Almost all SMIs in Malang Regency are grouped in the third or the lowest cluster. This lowest cluster is a cluster consisting of SMIs which are very vulnerable, due to their low performance and untested sustainability. This largest cluster requires great attention from various parties such as academia and especially the government. Aid projects, research, training, and mentoring are needed to encourage the development of SMIs at this stratum. This data classification can also be used to initiate SMIs communities. SMIs with similar classifications can form mutually supportive communities. These communities are important entities to support SMI innovation in developing their competitive advantage (Taneo, et al., 2019).

Currently, the data collection process for all SMI in Malang Regency is still being carried out by the Department of Industry and Trade of Malang Regency. This research can then be carried out again with a larger amount of data, so that a higher accuracy value can be obtained. Furthermore, further research is needed on the factors that cause the polarization of SMIs in the lowest cluster and how to resolve this problem.

REFERENCES

Kementerian Koperasi dan Usaha Kecil dan Menengah Republik Indonesia, "*Perkembangan Data Usaha Mikro, Kecil, Menengah (UMKM) dan Usaha Besar*". www.depkop.go.id. 2021.

Supriyanto, A., Basukianto, Rozaq, J.A., (2017), "*Klasterisasi UMKM dan Potensi Wilayah Berbasis Peta Sebagai Strategi Pengembangan Ekonomi Daerah,*" Jurnal Pekommas, 2(2), 143-150.

Akhmad, G.R., Susantiaji, A., (2020), "*Analisa Sebaran Klasifikasi Usaha Kecil Menengah (UKM) di Kabupaten Tegal,"* Geomedia, 18(1), 43-49.

Taneo, S.Y.M., Noya, S., Setiyati, E.A., Melany., (2019), "*Disruptive Innovation-Based Model of Sustainable Competitiveness Development in Small and Medium Food Industries,*" Matrik, 13(2), 153-163.

Madyaratry, L., Hadjomidjojo, H., Anggraeni, E., (2020), "*The Mapping of Sustainability Index in Small and Medium Enterprises: A Case Study in Lampung Indonesia,*" Jurnal Teknik Industri, 21(1), 58-69.

Kresnanto, N.C., (2020), "*Pemetaan Spasial Cluster Karakteristik UMKM Kabupaten Bantul*". Seminar Nasional Desiminasi Hasil Penelitian 2020 Universitas Janabadra Yogyakarta, 2020.

Raharjo, M.R., (2017), "*Analisis Algoritma Klasifikasi Dan Asosiasi Terhadap Atribut Data Pelaku Usaha Mikro Kecil dan Menengah (UMKM),*" Technologia, 8(3), 176-181.

Indriani, F., Kartini, D., (2018), "*Pola Klasifikasi Sektor Usaha UMKM dengan CART Menggunakan Seleksi Fitur Information Gain,*" Seminar Nasional Teknik Elektro dan Informatika (SNTEI).

Pech, M., Vrchota, J., (2020), "*Classification of Small- and Medium-Sized Enterprises Based on the Level of Industry 4.0 Implementation,*" Applied Science, 10(5150).

King, R.S., (2015), *Cluster Analysis and Data Mining: An Introduction*, Mercury Learning and Information: Boston.

Cohen, J., (1988), *Statistical Power Analysis for the Behavioral Science*. Laurence Erlbaum Associates: New York.

Taneo, S.Y.M., Noya, S., Setiyati, E.A., Melany., (2021), *Inovasi Disruptif: Strategi untuk Memenangkan Usaha*. Penerbit ANDI: Yogyakarta.

Venkatraman, N., Ramanujam, V., (1986), "*Measurement of Business Performance in Strategic Research,*" Academy of Management Review, 11(4), 801–814.