Application of the DBSCAN Algorithm in MSME Clustering using the Silhouette Coefficient Method

Mochammad Syahrul Abidin
Information Systems
Faculty of Engineering
Universitas Trunojoyo Madura
Bangkalan, Indonesia
syahrulabidin155@gmail.com

Yeni Kustiyahningsih
Information Systems
Faculty of Engineering
Universitas Trunojoyo Madura
Bangkalan, Indonesia
ykustiyahningish@trunojoyo.ac.id

Eza Rahmanita
Information Systems
Faculty of Engineering
Universitas Trunojoyo Madura
Bangkalan, Indonesia
eza.rahmanita@trunojoyo.ac.id

Budi Dwi Satoto
Information Systems
Faculty of Engineering
Universitas Trunojoyo Madura
Bangkalan, Indonesia
budids@trunojoyo.ac.id

Muhammad Iqbal Firmansyah
Information Systems
Faculty of Engineering
Universitas Trunojoyo Madura
Bangkalan, Indonesia
iqbalcode.nt@gmail.com

Abstract—MSMEs participate in the very important contribution of developing Indonesia's economy, where this industry contributes to GDP and also to the absorption of labor. Most MSMEs in Sidoarjo Regency are still constrained by financial management and the utilization of technology. This research will apply the DBSCAN method to clustering MSMEs in Sidoarjo for the purpose of finding patterns in characteristics related to capital, turnover, and workforce. The analysis will involve 1,479 MSMEs, while the research methodology applies the CRISP-DM method to guide the process from business understanding up to the implementation phase. Normalization using Simple Feature Scaling was applied before clustering. The results of this analysis provide insight that the best possible combination of the parameters in DBSCAN is epsilon (ε) 0.10 and MinPts 16, which gives the optimal value of Silhouette Score as 0.4304. It creates seven clusters, in which the third has the highest Silhouette value of 0.9326, indicating that there are high similarities recorded within that cluster. These results provide essential lessons to develop more targeted policy strategies and interventions for MSMEs in Sidoarjo and explore the capabilities of DBSCAN as an effective analytical tool in determining the characteristics of businesses in the region.

Keywords—MSMEs, DBSCAN, Clustering,Silhouette Coefficient

I. INTRODUCTION

MSMEs are categorized as the largest business group in Indonesia and are supposed to be the main basis of the national economy. Most MSMEs play a very important role in job creation and economic growth. It is therefore very important to provide the needed basis for maintaining stability in the economy amidst global uncertainty [1]. According to data from Kemenkop UKM, the amount of MSMEs in Indonesia has reached as big as 64.19 million business units. Contribution of MSMEs to the Indonesian economy is quite large, adding as much as 61.67% to the national GDP, or the equivalent of IDR 8,573.89 trillion. Besides, the MSMEs have a more significant contribution in labor absorption, accounting for about 97 percent of the total national workforce [2]. Whereas growth of the economy has persisted, the limited access to markets, capital, and technology continues to overburden MSMEs. Improvement in technology and, more importantly, data science, especially on clustering analysis, will allow the determination of MSME patterns and provide significant information relative to policies and strategies for development.

Sidoarjo Regency is one of the areas in Indonesia with abundant MSMEs. As one of the central industrial areas in East Java, Sidoarjo has high potential for MSME development. Even though this industrial area is known as an area that keeps growing,

©Asosiasi Prakarsa Indonesia Cerdas (APIC)

many MSMEs in Sidoarjo still fail to achieve optimal financial management and technology utilization [3]. The case of Sidoarjo therefore needs a better comparison in various characteristics of MSMEs, such as turnover, number of workers, and business scale, by local government and financial institutions among other stakeholders, since it consists of a large and diverse amount of MSMEs [4]. Overcoming this challenge is one of the approaches that can be applied using the clustering method or data grouping by CRISP-DM, which is very effective for finding solutions [5].

CRISP-DM is a methodology for data mining, dividing the process into six phases of business understanding, data understanding, data preparation, modeling, evaluation, and implementation. By using CRISP-DM, the whole research process-from understanding the business needs of MSMEs in Sidoarjo to implementing the results of the analysis for strategic decision-making support-can be given structure [6]. As one of the methods that can be done during the modelling process, the DBSCAN algorithm could be used [11][12][13]. The DBSCAN algorithm was first introduced in 1996 as a noisy clustering algorithm, which could handle variable shapes of data. This algorithm works by gathering data based on density, using core points, border points, and noise. The core points are formed when the data around them meet the minimum number of points required, while edge points are the data that are around the core point but the number of neighbors is less than the minimum number specified. Anything else that does not go into these two categories is what is called noise [7][14][15][16]. The characteristics of capital, monthly turnover, workforce, permits held, and marketplaces used can be used to group MSMEs by DBSCAN. These characteristics enable deeper insights into the development pattern of MSMEs in Sidoarjo Regency for data with noise or outliers.

Previous studies employed the DBSCAN algorithm to detect natural disaster patterns in Sumatra, a region high in tectonic activity and has geographical conditions prone to disasters. By taking advantage of the capability of DBSCAN for handling outliers and forming clusters without a pre-defined number, the current study managed to extract two optimal clusters by using the MinPts 4 and Eps 0.18 parameters. This result also gave a silhouette value of 0.46, depicting the precision of the clustering in the light of understanding disaster distribution for mitigation efforts within the region [8].

From the results above, DBSCAN becomes very powerful in clustering data with complex and high-noise characteristics in Sumatra. A similar approach can be replicated in similar contexts, say MSMEs in Sidoarjo, for the purposes of identifying patterns that could help stakeholders understand the characteristics of business variability and, hence, inform more targeted economic development strategies and policies. A DBSCAN-based analysis would be relevant not only to environmental studies and disaster mitigations but also to more targeted economic development strategies and policies. Such clustering enables the identification of MSMEs with high capital or turnover and also provides an overview of labor patterns, market access through markets, and the type of business licenses held by MSMEs in Sidoarjo. This will be very important, because each group of MSMEs requires a different policy strategy for its development.

©Asosiasi Prakarsa Indonesia Cerdas (APIC)

II. METHOD

A. Research Stages

Figure 1 depicts the research flow in this study.

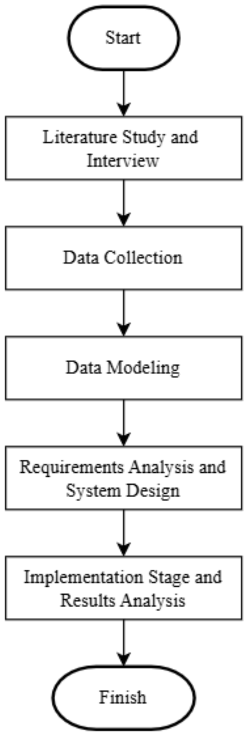

Figure 1. Research Flowchart

1. Literature Study and Interviews
   In order to ascertain the ideal number of clusters, a literature research and journal study are now being carried out to investigate ideas pertaining to data mining, clustering, data pre-processing, the DBSCAN method, the Simple Feature Scaling method, and the Silhouette Coefficient (SC). Additionally, data gathering and interviews were carried out with the Sidoarjo Regency's Cooperative and Micro Business Office.
2. Data collection
   - Data collection stage would then involve:
   - This present study used 1479 data of Batik MSMEs.
   - In this case, data that became an object of research is MSMEs within the area of Sidoarjo Regency.
   - Data source: Dinas Cooperatives and Micro Enterprises Service of Sidoarjo Regency.
   - Indicators of data from the service and the MSME actors are in consensus with Table 1.
3. Modeling Stage
   In the modeling stage of the cluster-based UMKM mapping application, the purpose and criteria used in this study are adapted to the needs of UMKM in Sidoarjo Regency. The pre-processing stage follows the modeling stage. Data normalization is applied through Simple Feature Scaling in this stage so that all data has a uniform scale before the clustering process. After that, the DBSCAN algorithm was used to form groups in data according to density. In order to find the best combination of

©Asosiasi Prakarsa Indonesia Cerdas (APIC)

epsilon (ε) and minimum number of points (MinPts) parameters, an evaluation is carried out using Silhouette Score at different values of the parameters. This is achieved by measuring Silhouette Score for several different combinations of ε and MinPts, then taking the highest-measured Silhouette Score as the best configuration parameter value. Optimal clustering thereafter, the result of the grouping is analyzed to give useful insights for the development of UMKM in Sidoarjo Regency.

4. Needs Analysis and System Design
   In this stage, the specifications for hardware and software requirements are stated, along with designing a system for MSME clustering. The hardware specifications will include:
   - Processor: AMD Athlon Silver 3050e
   - RAM: 4 GB
   - Storage: 256 GB Solid State Drive

   The following is the list of software specifications:
   - OS: Windows 11
   - Application: Google Colab, Microsoft Excel, StarUML

   Development of this software is expected to generate the following output:
   - The data on MSMEs comes from the Cooperatives and Micro Enterprises Service Sidoarjo Regency.
   - Suggestion results of MSME clustering with the application of the DBSCAN method.
   - Variable selection giving the biggest weight of influence on clustering.
   - Implementation, testing, and system analysis in order to get the right clustering result for working on the goals of clustering.
5. Implementation and Results Analysis Stage
   Implementation of the system model will be done regarding the design and needs analysis previously made. The process of clustering is expected to provide support for MSMEs and agencies in formulating appropriate recommendations on each indicator, so as to encourage the productivity of MSMEs in the Regency of Sidoarjo. Preprocessing at the trial stage normalizes the data using the Simple Feature Scaling method to have each variable constantly analyzed. Additionally, the DBSCAN method shall apply to group the data, which epsilon and minimum points are parameters determining through several experiments. The Silhouette Score method gives an idea about the quality of the clusters obtained and provides the best combination of the parameter values. DBSCAN clusters data points depending on their proximity using the eps and MinPts parameters without knowing the number of clusters in advance. The DBSCAN categorizes the data into core points and will automatically cluster the data in the radius of eps; data that does not meet these qualifications falls under the noise category. After clustering, the results obtained are analyzed for their quality, which in turn supports recommendations that can be developed for MSMEs.

B. DBSCAN Clustering

One of the known methods in clustering is DBSCAN, or Density-Based Spatial Clustering of Applications with Noise. Invented by Ester Martin, it emphasizes densities in space. Data groups with an arbitrary shape could be identified through this algorithm provided that certain levels of densities are detected [9]. The steps of the DBSCAN algorithm are as follows:

1. Initialize the parameters: MinPts and Eps
2. Choose a starting point, or point p, randomly.
3. For all points within the reachable density from this point to p, compute Eps value or distance using the Euclidean distance formula given by:

©Asosiasi Prakarsa Indonesia Cerdas (APIC)

$$d_{ij} = \sqrt{\sum_{a}^{p} \left(x_{ia} - x_{ja}\right)^2} \quad (1)$$

where x_ia is the a-th variable of object i (i=1, …, n; a=1, …, p) and d_ij is the Euclidean distance.

4. If the number of points within the radius of Eps is higher than MinPts, then p is identified as a core point, and a cluster with the basis of it is created.
5. Perform 3-4 for all the points: If p is the boundary and there is no other point which can be reached then move to the next point.

C. Silhoutte Coefficient

The Silhouette Coefficient is a method of evaluation that can provide a quality and strength measure of a cluster in clustering analysis. It gives an idea of how well an object is classified in a certain cluster because it takes into consideration the aspects of cohesion, such as the compactness inside the cluster and separation, which deals with the distance between clusters. A higher value of the Silhouette Coefficient guarantees that objects in a cluster have a good relationship among themselves and are well separated from other clusters [8]. Here are the main steps in the calculation of the silhouette coefficient:

1. The average distance of an observation, say i, to all other observations in a cluster is determined. It is termed as cohesion, calculated by the formula:

$$a(i) = \frac{1}{|A| - 1} \sum j \in_{A, j \neq i} d(i, j) \quad (2)$$

where j is any other observation in cluster A, and $d(i,j)$ is distance between observation i and j.

2. For observation i I calculate the average distance to all observations in the other clusters and choose the smallest as separation. This average is defined by:

$$d(i, C) = \frac{1}{|c|} \sum j \in d(i, j) \quad (3)$$

where $d(i,C)$ is the average distance of object i to all other objects in cluster C, and where C, and A is different from C.

3. Once the value of $d(i,C)$ has been calculated for every cluster C, choose the minimum through the expression:

$$b(i) = minc \neq Ad(i, C) \quad (4)$$

The cluster B yielding the smallest value of $d(i,C)$=$b(i)$ is called the nearest neighbor of object i and offers the best alternative cluster for the object.

4. For any observation i, the silhouette coefficient formula is given by:

$$S(i) = \frac{(B(i) - a(i)}{\max a(i), b(i)} \quad (5)$$

with:

- $S(i)$: the silhouette coefficient value for object i,
- $b(i)$: the minimum distance between object i and objects in other clusters C,

©Asosiasi Prakarsa Indonesia Cerdas (APIC)

- $a(i)$: the average distance between object i and all objects in the same cluster.

The value of Silhouette coefficient and its interpretation is shown in Table 1 [10].

Table 1. Silhouette coefficient values criteria

Silhouette coefficient value	Cluster structure
0.7 < SC ≤ 1	Cluster structure is very strong
0.5 < SC ≤ 0.7	Cluster structure is moderately strong
0.25 < SC ≤ 0.5	Cluster structure is weak
SC ≤ 0.25	Cluster is not structured

III. Result and Discussion

It also started the result and discussion of this study with dataset processing. The five criteria in the dataset are capital, monthly turnover, owned permit, marketplace, and workforce. The dataset used in this study contained 1,479 MSME data in Sidoarjo Regency. More information is shown in Table 2: MSME Dataset in Sidoarjo Regency.

Table 2. MSME Dataset

Business Name	Capital	Monthly Turnover	...	Permits Owned
7seven pomade	8000000	4000000	…	NIB
ab putra	500000000	5000000	…	NIB
abadi plastik mojo	10000000	2000000	…	SKU
...	...	...	…	…
...	...	...	...	...
zeger indonesia	50000000	4000000	…	NIB, IUMK
zico leather	5000000	15000000	…	NIB
ziva bubur bayi	50000000	18500000	…	NIB

After pre-processing the dataset, the time has come to start converting the MSME data listed in Table 3. This will concern changing the values of None or None to an empty string in the owned permit and marketplace columns, respectively. In the Owned Permit column, values for None and None are replaced by 0. For the Marketplace column, the values of None and None are replaced with 0. Additionally, after processing, the count of permits and count of marketplaces are calculated. If 0 is the value coming in the column, then it is still mentioned as 0 which later will be cast to float type 0.0. Because in the column owned permit, it counts the number of permits separated by commas and does the same in marketplace, calculating the total count.

Table 3. Numeric data conversion

Business Name	Capital	Monthly Turnover	...	Permits Owned
7seven pomade	8000000	4000000	…	1,0
ab putra	500000000	5000000	…	1,0
abadi plastik mojo	10000000	2000000	…	1,0
...	...	...	…	…
zeger indonesia	50000000	4000000	…	2,0
zico leather	5000000	15000000	…	1,0
ziva bubur bayi	50000000	18500000	…	1,0

©Asosiasi Prakarsa Indonesia Cerdas (APIC)

This entails normalizing and imputing the MSME dataset data using the simple feature scaling method before commencing any form of clustering. Numerical features at this stage include capital, turnover on a monthly basis, permits owned, marketplace, and workforce, all of which are normalized using the method. Table 4 displays the results of the pre-processing through the presentation of the MSME dataset after normalization.

Table 4. Data normalization

Business Name	Capital	Monthly Turnover	...	Labor
7seven pomade	0,072727	0,195122	…	0,285714
ab putra	1,000000	0,243902	…	0,571429
abadi plastik mojo	0,090909	0,097561	…	0,285714
...	...	...	…	…
...	...	...	...	...
zeger indonesia	0,454545	0,195122		0,571429
zico leather	0,045455	0,731707	…	1,000000
ziva bubur bayi	0,454545	0,902439	…	1,000000

After normalization, several clustering experiments were performed using the DBSCAN algorithm to find the best number of clusters. Each experiment has different values for Eps and MinPts. In this paper, the values of Eps range from 0.10 to 0.14, while MinPts values range from 12 to 16. The result of clustering can be found in the following table:

Table 5. Data cluster results

Experiment	Eps	MinPts	Number of Clusters	Noise	Averange Silhouette score
1	0,10	12	12	262	0,4168
2	0,10	13	11	280	0,4208
3	0,10	14	8	322	0,4295
4	0,10	15	7	336	0,4283
5	0,10	16	7	339	0,4304
6	0,11	12	12	219	0,3984
7	0,11	13	13	224	0,4006
8	0,11	14	9	292	0,4042
9	0,11	15	8	312	0,4024
10	0,11	16	7	325	0,4241
11	0,12	12	13	200	0,3973
12	0,12	13	12	218	0,3988
13	0,12	14	9	278	0,3963
14	0,12	15	9	285	0,3963
15	0,12	16	7	313	0,4179
16	0,13	12	12	194	0,3996
17	0,13	13	12	197	0,4001
18	0,13	14	9	240	0,3994
19	0,13	15	9	263	0,3914
20	0,13	16	7	305	0,4145
21	0,14	12	12	164	0,3868
22	0,14	13	12	166	0,3869

©Asosiasi Prakarsa Indonesia Cerdas (APIC)

23	0,14	14	10	202	0,3924
24	0,14	15	11	203	0,3797
25	0,14	16	10	220	0,3807

According to the experimental result in Table 5, the best Eps of the optimal clustering can be set as 0.10 and MinPts at 16, since this setting gives out the highest silhouette coefficient, namely 0.4304. With an increase in the value of MinPts, cluster quality tends to be improved; after a certain point, however, cluster quality can worsen. The higher the value of MinPts is, the less likely it is that a cluster forms. The value of the minimum number of objects which have to be adjacent in a cluster becomes higher. It would need more objects to form one cluster. In Table 6, the result table for clustering using DBSCAN displays optimal Eps and MinPts parameter values.

Table 6. Clustering results 7 clusters

Business Name	Capital	Monthly Turnover	...	Cluster
angsle ronde	0,027273	0,243902	…	C1
anies craft	0,009091	1,000000	…	C4
animo laundry	0,090909	0,146341	…	C1
...	...	...	…	…
...	...	...	…	…
zalfa wijaya	0,027273	0,048780	…	C0
zarryz collction	0,454545	0,487805	…	C1
zeger indonesia	0,454545	0,195122	…	C1

Table 6: The clustering result of MSME data into 7 clusters. Cluster -1 contains 339 MSMEs ungrouped into a particular cluster, so it is regarded as noise. While the cluster that has the largest number of businesses is Cluster 0 with 693 businesses showing similarities because the dominant characteristics are similar, such as similar capital and labor. Cluster 1, on the contrary, presents 307 enterprises with some features, too. Further still, Cluster 2 comprises 48 enterprises, while the smaller ones include Cluster 3 to Cluster 6, each comprising about 23-24 enterprises. These clusters present more specific attribute variations and are confined to a small group of MSMEs. These smaller membership clusters can describe unique characteristics and thus may have different approaches to development for each type of business.

Table 7. Silhouette Score value of each cluster

Cluster	Silhouette Coefficient Value
C0	0,4189
C1	0,2695
C2	0,9131
C3	0,9326
C4	0,8747
C5	0,7057
C6	0,6877

Table 7: Silhouette Coefficient Value for each cluster in this dataset, where it depicts the measure of how well MSMEs in this dataset have been grouped. Cluster C3, with a value of 0.9326, gives very good results, meaning businesses in this cluster are quite similar to each other and almost detached from other clusters. This value is included in

©Asosiasi Prakarsa Indonesia Cerdas (APIC)

the category "very strong cluster structure" according to the criteria in Table 1. Figure 2 shows a graph of the distribution of turnover in accordance with the type of business.

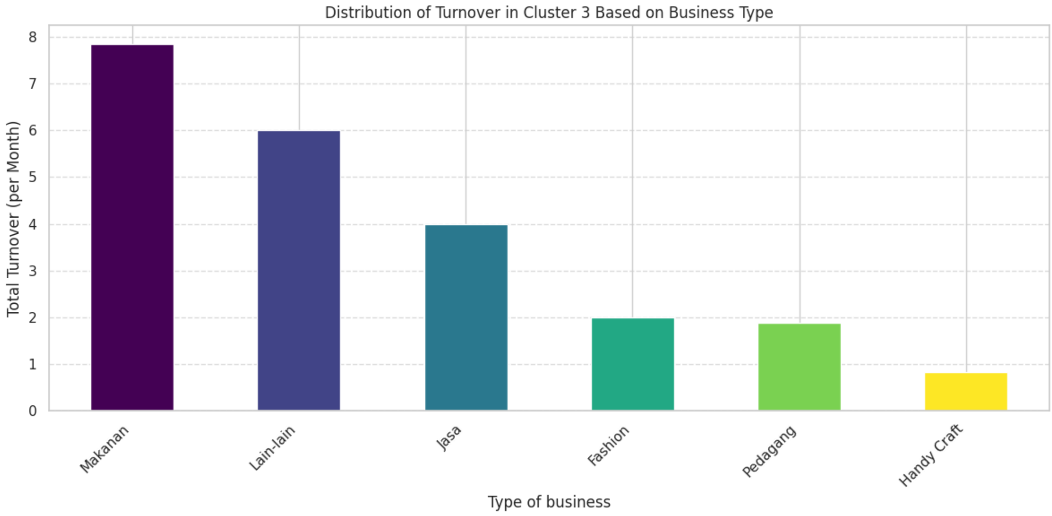

Figure 2. Distribution graph of turnover in Cluster 3

Figure 2: Distribution of average turnover for different types of businesses in cluster 3 of the MSMEs grouping in Sidoarjo Regency carried out using the DBSCAN method. This graph compares the different types of businesses by their average turnover per month. The Food business type has the highest average turnover, which means this sector contributes a great deal to the cluster. This is followed by the Miscellaneous business type, then the Service business type, while third place is taken by the Fashion sector. Below it, there is the Trader sector, while the Handy Craft sector is with the lowest average of turnover. From this graph, it can be observed that food dominates in the cluster's turnover, while the Handy Craft is the least contributor. Identifying this cluster is targeted at finding the pattern of the turnover that different kinds of MSMEs have demonstrated in Sidoarjo and defining those business groups that bear similar characteristics. The business data can be clustered with the help of the DBSCAN method based on the density of data. In such a way, groups having similar turnovers can be identified. It is envisaged that this kind of analysis will provide the basis for more appropriate decision-making, offering assistance or policy interventions that meet better the needs of each cluster and development of strategies related to increasing the turnover within those sectors that have the highest potential.

IV. CONCLUSION

It can be concluded that DBSCAN, or Density-Based Spatial Clustering of Applications with Noise, is the method applied in clustering MSME data in Sidoarjo Regency. This clustering is done to identify the pattern and cluster on the basis of similar characteristics by considering some variables like capital, monthly turnover, number of workers, marketplaces used, and business licenses. Here, the best combination of parameters is given by an eps value of 0.10 and minPts of 16, with the optimal value of Silhouette score being 0.4304. Clustering results have formed 7 clusters. The best Silhouette score value comes out to be in the 3rd cluster, with a value of 0.9326. These results have shown that the selection of appropriate parameters by the DBSCAN algorithm is a very influential issue regarding the identification of optimal cluster patterns in conformity with the characteristics of the MSME data in existence. The DBSCAN method could be used for the region as an important means in assessing features and distribution of MSMEs and providing concrete data for more efficient business strategies.

©Asosiasi Prakarsa Indonesia Cerdas (APIC)

REFERENCES

[1] S. R. R. S. R. P. Djoko Wahyudi, "Empowerment Of Micro Small And Medium Enterprises A Presto Milkfish In The Era Of Pandemics Kanigaran Sub District Probolinggo City," *INTEGRITAS : Jurnal Pengabdian,* vol. 5, pp. 26 - 32, 2021.

[2] A. R. L. S. Maria Mu'ti Wulandari, "Efforts To Development Of Entrepreneurship Koperasi And Micro, Small And Medium Business Post Covid-19," *Soedirman Law Review,* vol. 5, pp. 120 - 136, 2023.

[3] F. G. S. G. M. K. Fanniar Aurelia, "Peranan Sistem Digital Accounting Terhadap Perkembangan Bisnis," *Jurnal Manajemen dan Ekonomi Bisnis,* vol. 3, pp. 26 - 40, 2023.

[4] K. M. B. S. A. F. Indah Permata Sari, "Analisa Kebijakan Publik Terhadap UMKM," *Jurnal Ilmiah Mahasiswa Pendidikan Sejarah,* vol. 8, pp. 552 - 559, 2023.

[5] L. E. Z. O. H. H. Dwi Utari Iswavigra, "Marketing Strategy UMKM dengan CRISP-DM Clustering dan Promotion Mix Menggunakan Metode K-Medoids," *Jurnal Informasi dan Teknologi,* vol. 5, pp. 45 - 54, 2023.

[6] A. R. I. A. F. Dwi Astuti, "Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering," *Journal of Informatics, Information System, Software Engineering and Applications,* vol. 1, pp. 60 - 72, 2019.

[7] D. A. W. W. Zicola Vladimir, "Klasterisasi Topik Skripsi Informatika Dengan Metode DBSCAN," *Jurnal Algoritme* , vol. 3, pp. 82 - 90, 2022.

[8] M. A. H. R. W. Aulia Fadhli Boer, "Pengelompokan Daerah Rawan Bencana di Pulau Sumatera dengan," *Prosiding Seminar Nasional,* vol. 6, pp. 389 - 400, 2023.

[9] M. Nana Nurhaliza, "Clustering of Data Covid-19 Cases in the World Using DBSCAN Algorithms," *Indonesian Journal of Informatic Research and Software Engineering,* vol. 1, pp. 1 - 8, 2021.

[10] A. Z. A. H. P. L. I. Rahmatina Hidayati, "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering," *Techno.Com,* vol. 20, pp. 186 - 197, 2021.

[11] Y. L. W. L. J. H. Zhichao Li, "Crowdsourcing Logistics Pricing Optimization Model Based on DBSCAN Clustering Algorithm," *IEEE Access,* vol. 8, pp. 92615 - 92626, 2020.

[12] K. J. S. R. M. Z. Shapol M. Mohammed, "Glove Word Embedding and DBSCAN algorithms for Semantic Document Clustering," *IEEE Access,* pp. 211 - 216, 2021.

[13] D. Deng, "Research on Anomaly Detection Method Based on DBSCAN Clustering Algorithm," *International Conference on Information Science, Computer Technology and Transportation,* pp. 439 - 442, 2021.

[14] S.-S. Li, "An Improved DBSCAN Algorithm Based on the Neighbor Similarity and Fast Nearest Neighbor Query," *IEEE Access,* vol. 8, pp. 47468 - 47476, 2020.

[15] G. Y. Adil Abdu Bushra, "Comparative Analysis Review of Pioneering DBSCAN and Successive Density-Based Clustering Algorithms," *IEEE Access,* vol. 9, pp. 87918 - 87935, 2021.

©Asosiasi Prakarsa Indonesia Cerdas (APIC)

[16] D. Deng, “DBSCAN Clustering Algorithm Based on Density,” *International Forum on Electrical Engineering and Automation,* pp. 949 - 953, 2021.

[17] W. H. X. X. J. Y. Yong Huang, “An empirical study of personalized advertising recommendation based on DBSCAN clustering of sina weibo user-generated content,” *10th International Conference of Information and Communication Technology,* pp. 303 - 310, 2021.

[18] J. S.-P. Juan Carlos PERAFA´ N-LO´ PEZ, “An unsupervised pattern recognition methodology based on factor analysis and a genetic-DBSCAN algorithm to infer operational conditions from strain measurements in structural applications,” *Chinese Journal of Aeronautics,* pp. 165 - 181, 2021.

[19] N. H. R. C. d. A. Stiphen Chowdhurya, “Feature weighting in DBSCAN using reverse nearest neighbours,” *Elsevier,* pp. 1 - 15, 2023.

[20] A. A. S. G. F. L. G. Kevin Surya Pranata, “Development clustering system IDX company with k-means algorithm and DBSCAN based on fundamental indicator and ESG,” *7th International Conference on Computer Science and Computational Intelligence,* pp. 319 - 327, 2023.

[21] C. Y. A. L. Y. C. L. C. Bing Ma, “A Faster DBSCAN Algorithm Based on Self-Adaptive Determination of Parameters,” *10th International Conference on Information Technology and Quantitative Management,* pp. 113 - 120, 2023.

©Asosiasi Prakarsa Indonesia Cerdas (APIC)