KOMPUTASI: JURNAL ILMIAH ILMU KOMPUTER DAN MATEMATIKA
Vol. 20 (1) (2023), 56-70, p-ISSN: 1693-7554, e-ISSN:2654-3990

Distribution Analysis Active Small and Medium Industries Bogor City Using K-means Clustering

Siska Andriani[1], **Faradilla Rizqiyyah**[2], **Rahmat H A Asir**[3]

[1,2]Department of Computer Science, Faculty of Mathematics and Natural Science, Pakuan University, Bogor, West Java, 16143, Indonesia
[3] Office of Cooperatives, Small and Medium Enterprises, Bogor City Trade and Industry, Bogor, West Java, Indonesia

Abstract

Abstract Small and Medium Industries (IKM) is one sector that contributes to driving economic growth, one of which is in West Java province, Bogor city. The number of active IKM in the city of Bogor in 2021 based on the survey results was 1,189, while in 2022 the number of small and medium industries (IKM) active in the city of Bogor based on the survey results was 1,766. The purpose of this study was conducted to determine the distribution of active small and medium industries (IKM) in Bogor city. So, this research can provide solutions related to the government or agencies to assist in building and developing IKM. In this research, the method used is the Knowledge Discovery in Database method, Where the stages are data selection, pre-processing, transformation, data mining and evaluation. Determination of the number of clusters is done using the elbow method. After determining with Elbow, the data will be represented using k-means clustering. The results of the k-means clustering algorithm yield 3 clusters, with each cluster 0 criterion being the distribution of low IKM with a total of 45 sub-districts. Cluster 1 is the distribution of medium IKM with the number of sub-districts is 14, and cluster 2 is the distribution of high IKM with the number of sub-districts totaling 9. The evaluation in this study used the silhouette coefficient method, from the data used it produced a cluster value of 0.56 which means that it is included in the clustering criteria with a good structure..

Keywords: *Data Mining; Clustering; K-Means; Elbow; Silhouette Coefficient*

1. Introduction

Small and Medium Industries (IKM) is one sector that contributes to driving economic growth, one of which is in West Java province, Bogor city. The number of active small and medium industries (IKM) in the city of Bogor in 2021 based on the survey results was 1,189, while in 2022 the number of small and medium industries (IKM) active in the city of Bogor based on the survey results was 1,766. This proves that the role of small and medium industries (IKM) provides aspects in building the community's economy, namely as job creation and the key to economic growth.

Even though IKM provides an important role, in developing a business, the government still has obstacles to assist it, such as providing assistance for the development of IKM by intervening

*Corresponding author. *E-mail address: siska.andriani@unpak.ac.id*
Received: 05 November 2022, Accepted: 25 December 2022 and available online 31 January 2022
DOI: https://doi.org/10.33751/komputasi.v20i1.6559

randomly on Instagram. Therefore, it is necessary to group active small and medium industries (IKM) in the city of Bogor to make it easier for agencies to help carry out the development of IKM. The research related to the analysis of the distribution of active small and medium industries (IKM) in the city of Bogor using the k-means clustering method was carried out by [1]. In this research, the UMKM grouping in South Sumatra Regency was carried out by applying the k-means clustering algorithm.

The stages of the research carried out include the business understanding phase, data understanding phase, data processing phase, Modeling Phase, evaluation phase and dissimination phase. In testing 15 business data, this study succeeded in applying the k-means clustering algorithm to identify and classify UMKM with test results of 53% of data to cluster 1 of 8 data, 40% of data to cluster 2 of 6 data and 7% of data to cluster 3 as much as 1 data. The results of this calculation have also been tested using rapid miner software and produce the same data.

The results of the calculation of the K-means clustering algorithm can be used as input for stakeholders in the assistance, development and management of existing UMKM, such as the Office of Cooperatives and UMKM or other relevant agencies. with this. With the hope that existing UMKM can develop better.

Based on the previous research above, it can be concluded that to analyze the distribution of active small and medium industries (IKM) in Bogor city, it can be solved by computer science techniques, namely the implementation of data mining using the k-means clustering method.

1.1. Data Mining

Data mining is a combination of a number of computer science disciplines that are defined as the process of discovering new patterns from very large data sets, including methods that are slices of artificial intelligence, machine learning, statistics, and database systems [2]. Based on the understanding of data mining according to the experts mentioned, it can be concluded that data mining is a process of searching data automatically to obtain a model from a large database. Data Mining is a method for finding hidden information in databases and part of the Knowledge Discovery in Databases (KDD) process for finding useful information and patterns in data. Overall, the Knowledge Discovery in Database (KDD) process can be described as follows [3]:

1. Data Selection
   Data selection from a set of operational data needs to be done before the information mining stage in KDD begins. The selected data will be used for the data mining process.
2. Pre-processing/cleaning (data cleaning)
   Before the data mining process is carried out, it is necessary to carry out a cleaning process on the data that is the focus of KDD. The cleaning process includes removing duplicate data, checking inconsistent data, and correcting errors in data.
3. Transformation
   Transformation is changing data into a form suitable for mining.
4. Data Mining
   Data mining is the process of looking for patterns or interesting information in selected data using certain techniques or methods.
5. Interpretation/Evaluation
   Patterns of information resulting from the data mining process need to be displayed in a form that is easily understood by interested parties. Such as using a visualization or display that can explain the output of the system.

1.2. Elbow Method

Elbow is a method used to generate information in determining the best number of clusters by looking at the percentage comparison between the number of clusters that will form an elbow

at a point. The Elbow method is a method used to generate information in determining the best number of clusters by looking at the percentage comparison of the number of clusters that will form an elbow at a point. This method provides ideas by selecting cluster values and then adding these cluster values to be used as a data model in determining the best cluster. And besides is the percentage of the resulting calculation a comparison between the number of clusters added [4]. The results of the different percentages of each cluster value can be shown using graphics as a source of information. If the first cluster value with the second cluster value gives a corner in the graph or the value has the largest decrease, then the cluster value is the best. To get a comparison is to calculate the SSE (Sum of Square Error) of each cluster value.

This method and analysis is used for selecting the optimal number of clusters or groups. In the following, the elbow algorithm is presented in determining the number of groups formed) [5]. That is based on the sum of square error (SSE). Because the greater the number of K clusters, the smaller the SSE value. The following are the stages of the Elbow method algorithm in determining the value of k in K-Means:

1. Initialize the initial value k;
2. Increase the value of k;
3. Calculate the sum of the square errors for each k value;
4. Analysis of the results of the sum of square error from the value of k which has decreased drastically;
5. Find and set the angled k value.

In the Elbow method the best cluster value will be taken from the Sum of Square Error (SSE) value which has a significant decrease and is in the shape of an elbow. To calculate SSE using the formula.

$$SSE = \Sigma_{k=1}^{k} \Sigma_{x1_g sk} ||X_i - C_k||^2 \tag{1}$$

Where K is the number of groups used in the K-means algorithm Xi is the amount of data and Ck is the number of clusters in the kth cluster.

1.3. K-Means Algorithm

The k-means algorithm is an algorithm used for clustering and this algorithm is in a non-hierarchical form which has a relatively fast computation time [6]. According to another opinion, k-means is a non-hierarchical data clustering method that seeks to partition existing data into the form of one or more clusters or groups so that data that has the same characteristics are grouped into the same cluster and data that has the same characteristics. different groups are grouped into other groups [7]. The K-Means algorithm is an algorithm that groups data by trying to separate data into groups so that data that has similarities is in the same group. However, different data are classified in other groups [8].

This algorithm will determine how many clusters to use, then determine the value for each cluster. Next, calculating calculates the distance of each data with a predetermined initial value. The data is then placed in the nearest cluster. Clusters that already have data are then calculated for the average value of each cluster which will later be used for the new initial value in calculating the distance from each data. The iteration continues until the new cluster value is the same as the previous cluster value or does not change [9]. The purpose of k-means is grouping data by maximizing the similarity of data in one cluster [10]. In the following, to calculate the distance to the center of the cluster, use the formula in equation (2).

$$D(i,j) = \sqrt{(X1_i - X1_j)^2 + ... + (Xk_i - Xk_j)^2} \tag{2}$$

1.4. Silhouette Coefficient Method

The silhouette coefficient method serves to test the quality of the resulting clusters as well as a method for validating a cluster that combines the cohesion method and the separation method [11]. To calculate the value of the silhoutte coefficient, it takes the value of the distance between objects using the Euclidean distance method. The stages in determining the value of the silhouette coefficient are as follows:

1. For each object i the average value of one point is calculated with all objects in one cluster. Then an average value called ai will be obtained.
2. For each object i the minimum value of the average distance from one point to another is calculated in a different cluster. Then a minimum average value called bi will be obtained.
3. Then after all the values are known, the value of the silhouette coefficient can be determined using the following formula in equation (3)

$$S_i = \frac{b_i - ai}{max(a_i - b_i)} \tag{3}$$

S_i : silhouette coefficient value.
a_i : the average distance from one point to all data in onenclusters.
b_i : minimum average distance from one point to another in a different cluster.

Subjective criteria for cluster measurements on the silhouette coefficient can be seen in the following table.

Table 1. Clustering value in the silhotte Coefficient Method

Silhoutte Coefficient Value	Criteria
0.71 - 1.00	Strong Structure
0.51 - 0.70	Good Structure
0.26 - 0.50	Weak Structure
$\leq$ 0,25 Poor	poor Structure

2. Method

In this research, the method used is the Knowledge Discovery in Database method. This method is a method used in data mining. The following is the flow of the Knowledge Discovery in Database (KDD) research method. Where the stages are data selection, pre-processing, transformation, data mining and evaluation. The stages can be seen in Figure 1.

2.1. Selection

The initial stages carried out were data search, data collection, and data selection which were carried out at the Office of Cooperatives, Small and Medium Enterprises, Trade and Industry in the City of Bogor. The data taken is in the form of active small and medium industry (IKM) data for the city of Bogor in 2021 – 2022 with an amount of data of 2,393 data. The data to be used in this study is data on active small and medium industries (IKM) based on sub-districts in Bogor city of 68 sub-districts. The data used for research has obtained permission from the agency. The following is data on active small and medium industries (IKM) based on sub-districts in the city of Bogor in 2021 - 2022.

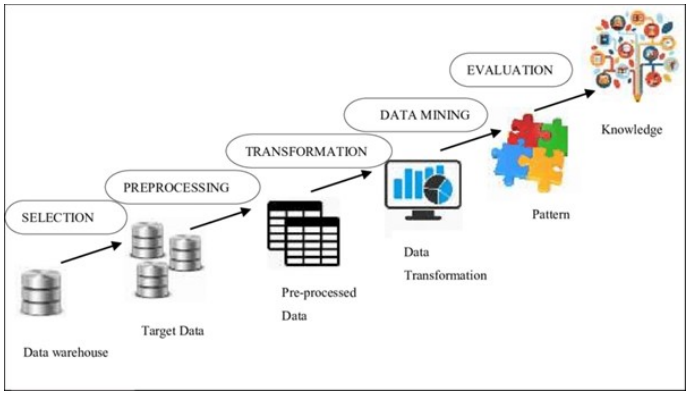

Figure 1. The stages of KDD

2.2. Pre-Processing

In the pre-processing stage, the attribute format will be changed to a nominal data type on the village attribute. The attributes that serve as a reference for the placement of the clustering results are the attributes of the Number of IKM 2021 and the Number of IKM 2022 because these attributes are of integer or number data type.

2.3. Transformation

The transformation stage in the research is the stage of adjusting the data to suit the needs in processing data mining applications. At this stage data scaling and normalization is carried out which aims to facilitate data processing at a later stage and equalize the data range so that all existing data is balanced.

2.4. Data Mining

Data mining is a process for finding interesting information on data using certain methods. One of the methods used in this study is using the clustering method with the k-means algorithm. K-means is one of the clustering algorithms used to group data into several groups with several clusters. The K-means method groups existing data into several groups, for example, data in one group has the same characteristics as one another and data with different characteristics will be separated into different groups. Where each cluster has a center point called the centroid. The following are the steps for performing calculations using the K-means algorithm, as follows:
a. Choose the desired number of clusters (k) in the dataset.
b. Determine the center point (centroid) randomly at the initial stage
c. Calculates the closest distance of each data to the centroid. To calculate the closest distance to the center point (centroid) use the Euclidean distance formula.
d. Recalculate the cluster center with the new cluster membership. Cluster center is the average of all data in a cluster.

3. Result and Discussion

The results of this study are the grouping of active small and medium industries (IKM) based on the Bogor urban village from 2021 – 2022 using the python programming language with Google colaboratory. With this grouping, related agencies can assist in building and developing IKM.

3.1. Selection

The data selection stage was carried out at the Bogor City Office of Cooperatives, Small and Medium Enterprises, Trade and Industry. The data taken is in the form of active IKM data for

the city of Bogor based on sub-districts with a total of 68 sub-districts and a total of 2,393 data. Where, 2,393 data is the combined number of 2021-2022 IKM, 2021 IKM totaling 627 IKM and 2022 IKM totaling 1,766 IKM. The following is the active IKM data for the city of Bogor by sub-district.

Table 2. IKM Data of Kelurahan Bogor City 2021 – 2022

No	Kelurahan	Number of IKM 2021	Number of IKM 2022
1	Balumbang Jaya	2	10
2	Bubulak	7	16
3	Cilendek Barat	8	23
4	Cilendek timur	8	19
5	Curug	8	19
6	Curug Mekar	13	13
7	Gunung batu	8	15
8	Loji	4	20
9	Margajaya	4	11
10	Menteng	10	20
11	Pasir jaya	2	25
12	pasir Kuda	22	38
13	Pasir Mulya	4	14
14	Semplak	5	10
15	Sindangbarang	15	33
16	Situ Gede	8	18
17	Batu Tulis	10	54
18	Bojong Kota	9	44
10	Bondongan	7	30
20	Cikaret	9	63
21	Cipaku	8	52
22	Empang	17	27
23	Genteng	1	18
24	Harjasari	1	5
25	Kertamaya	5	14
26	Lawang Gintung	4	10
27	Muarasari	3	18
28	Mulyaharja	5	84
29	Pakuan	2	7
30	Pamoyanan	12	27
31	Rancamaya	7	7
32	Ranggamekar	4	48
33	Babakan	10	14
34	Babakan Pasar	1	6
35	Cibogor	7	9
36	Ciwaringin	6	9
37	Gudang	7	14
38	Kebon Kelapa	1	22
39	Pabaton	5	9
40	Paledang	4	7
41	Panaragan	2	5
42	Sempur	5	14
43	Tegallega	9	16
44	Baranangsiang	14	33
45	Katulampa	23	62

No	Kelurahan	Number of IKM 2021	Number of IKM 2022
46	Sindangrasa	8	17
47	Sindangsari	6	30
48	Sukasari	9	14
49	Tajur	11	19
50	Bantar Jati	19	41
51	Cibuluh	14	42
52	Ciluar	6	46
53	Cimahpar	6	24
54	Cipagiri	19	41
55	Kedung Halang	11	47
55	Kedung Halang	11	47
56	Tanah Baru	20	48
57	Tegal Gundil	19	40
58	Cibadak	16	27
59	Kayumanis	4	54
60	Kebon Pedes	26	31
61	Kedung Badak	29	48
62	Kedung Jaya	13	16
63	Kedung Waringin	17	50
64	Kencana	2	8
65	Mekarwangi	6	27
66	Sukadamai	12	23
67	Sukaresmi	6	19
68	Tanah sereal	12	21

3.2. Pre-Processing

The preprocessing carried out in this study is to determine the names of the kelurahan which are recognized as objects to be clustered (by changing the attribute format to a nominal data type) and the case attributes for the number of IKM 2021 and the attribute for the number of IKM 2022 which will be used as a reference for placement. clustering results (integer data type for numbers). The results of data pre-processing that has been carried out at the Google collaboratory using the Python programming language are shown in Figure 2.

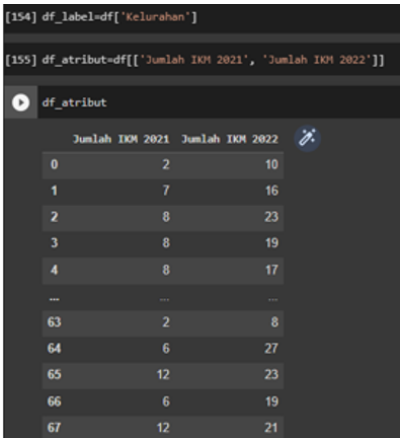

Figure 2. Pre-processing Process

3.3. Transformation

at this stage data changes are made to facilitate processing in accordance with the data mining process. Data changes made to. The attributes of the number of IKM 2021 and the number of IKM 2022 data are converted into data normalization forms. The transformation stage in this study was carried out by normalizing data on active IKM data in the city of Bogor. The following is the result of the transformation process.

Table 3. Results of the Transformation Process

Index	Number of IKM 2021	Number oh IKM 2022
0	-1,126605826	-0,94925573
1	-0,330252435	-0,592629268
2	-0,170981757	-0,176565062
3	-0,170981757	-0,414316037
4	-0,170981757	-0,533191524
5	0,625371633	-0,770942499
6	-0,170981757	-0,652067012
7	-0,808064469	-0,354878293
8	-0,808064469	-0,889817987
9	0,147559599	-0,354878293
10	-1,126605826	-0,057689575
11	2,058807735	0,715001093
12	-0,808064469	-0,711504756
13	-0,648793791	-0,94925573
14	0,943912989	0,417812375
15	-0,170981757	-0,473753781
16	0,147559599	1,666004993
17	-0,011711079	1,071627556
18	-0,330252435	0,239499144
19	-0,011711079	2,200944686
20	-0,170981757	1,547129505
21	1,262454345	0,061185913
22	-1,285876504	-0,473753781
23	-1,285876504	-1,246444449
24	-0,648793791	-0,711504756
25	-0,808064469	-0,94925573
26	0,967335148	-0,473753781
27	-0,648793791	3,449137304
28	-1,126605826	-1,127568961
29	0,466100955	0,061185913
30	-0,330252435	-1,127568961
31	-0,808064469	1,30937853
32	0,147559599	-0,711504756
33	-1,285876504	-1,187006705
34	-0,330252435	-1,008693474
35	-0,489523113	-1,008693474
36	-0,330252435	-0,711504756
37	-1,285876504	-0,236002806
38	-0,648793791	-1,008693474
39	-0,808064469	-1,127568961
40	-1,126605826	-1,246444449
41	-0,648793791	-0,711504756
42	-0,011711079	-0,592629268
43	0,784642311	0,417812375

Index	Number of IKM 2021	Number oh IKM 2022
44	2,218078413	2,141506942
45	-0,170981757	-0,533191524
46	-0,489523113	0,239499144
47	-0,011711079	-0,533191524
48	0,306830277	-0,414316037
49	1,580995701	0,893314325
50	0,784642311	0,952752068
51	-0,489523113	1,190503043
52	-0,489523113	-0,117127318
53	1,580995701	0,893314325
54	1,580995701	1,249940787
55	1,740266379	1,30937853
55	1,740266379	1,30937853
56	1,580995701	0,833876581
57	1,103183667	0,061185913
58	-0,808064469	1,666004993
59	2,695890447	0,298936887
60	3,173702482	1,30937853
61	0,625371633	-0,592629268
62	1,262454345	1,428254018
63	-1,126605826	-1,068131218
64	-0,489523113	0,061185913
65	0,466100955	-0,176565062
66	-0,489523113	-0,414316037
67	0,466100955	-0,29544055

3.4. Data Mining

At this stage, data processing will be carried out using the clustering method with the k-means algorithm. Data processing is processed using the Python programming language with the help of Google colaboratory software. In using this method, the initial stage will be to determine the number of clusters to be formed. Determination of the number of clusters is done using the elbow method. This method is used because it determines the optimal k value in the clustering method. Determining the Number of Clusters Using the Elbow Method is shown in Figure 3.

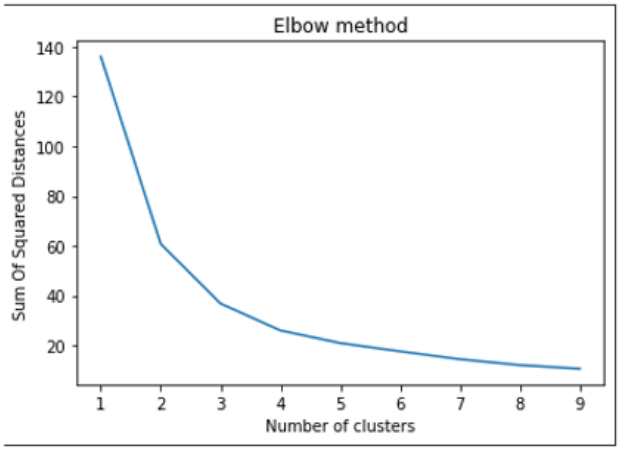

Figure 3. The results of determining the number of clusters using the Elbow method are shown **in**

The stage of determining the number of clusters has been obtained by the elbow method. So, in this study the clusters formed were 3 clusters, cluster 0 showing a low distribution of IKM,

cluster 1 showing a moderate distribution of IKM, cluster 2 showing a high distribution of IKM. After determining the number of clusters using the elbow method, the data that has been done in the previous process will be represented using k-means clustering in the form of k-means.fit. The number of k that will be declared in the process is 3 clusters, because previously it has been determined using the elbow method. The result of the process using the k-means algorithm shown in Figure 4.

```
[67] k_means= KMeans(n_clusters=3, random_state=123)

[68] k_means.fit(std_atribut)
     print(k_means)

     KMeans(n_clusters=3, random_state=123)
```

Figure 4. The result of the process using the k-means algorithm

The output results above have shown the number of clusters to be formed in this study as many as 3 clusters. In the next stage, a process will be carried out to find out the centroid value formed by the k-means algorithm. The results of the centroid values shown in Figure 5.

```
[88] print(k_means.cluster_centers_)

     [[-0.45059028 -0.60187514]
      [ 1.62650161  0.83812213]
      [-0.27716221  1.70563016]]
```

Figure 5. The centroid values

After getting the centroid value that was successfully formed in the process, clustering or grouping can be done by calculating the distance of each data from the centroid. After the calculation process is carried out, the iteration results are obtained 3 times. After knowing the centroid value and knowing the number of iterations obtained in the previous process. Thus, the grouping of IKM in Bogor city sub-districts can be produced, by producing 3 clusters.

Cluster 0 is a category of low IKM distribution grouping, which means that the sub-districts included in cluster 0 have a small number of IKM, so agencies or the government need to improve in building and developing IKM in these sub-districts. Improvements made such as how to encourage people in these sub-districts to be interested in building and starting businesses, conducting training to build and make good and right businesses.

Cluster 1 is a medium-sized IKM distribution grouping category, which means that the sub-districts included in cluster 1 have a moderate number of IKM. So it is necessary to do development in increasing IKM in the sub-districts that are included in the cluster. The development carried out is a way to increase product sales on social media, a way to manage good finances in running a business.

Cluster 2 is a grouping category for the distribution of high IKM, which means that the sub-districts included in cluster 2 have a large number of IKM. Kelurahan which has a large number of IKM means it can contribute to boosting economic growth in the city of Bogor.

The following is the result of clustering that has been processed using the Python programming language with the help of the Google colaboratory.

Table 4. Clustering Results

No	Kelurahan	*Cluster*	Information
1	Balumbang Jaya	0	Rendah
2	Bubulak	0	Rendah
3	Cilendek Barat	0	Rendah
4	Cilendek Timur	0	Rendah
5	Curug	0	Rendah
6	Curug Mekar	0	Rendah
7	Gunung Batu	0	Rendah
8	Loji	0	Rendah
9	Margajaya	0	Rendah
10	Menteng	0	Rendah
11	Pasir Jaya	0	Rendah
12	Pasir Kuda	1	Sedang
13	Pasir Mulya	0	Rendah
14	Semplak	0	Rendah
15	Sindangbarang	1	Sedang
16	Situ Gede	0	Rendah
17	Batutulis	2	Tinggi
18	Bojong Kerta	2	Tinggi
19	Bondongan	0	Rendah
20	Cikaret	2	Tinggi
21	Cipaku	2	Tinggi
22	Empang	1	Sedang
23	Genteng	0	Rendah
24	Harjasari	0	Rendah
25	Kertamaya	0	Rendah
26	Lawang Gintung	0	Rendah
27	Muarasari	0	Rendah
28	Mulyaharja	2	Tinggi
29	Pakuan	0	Rendah
30	Pamoyanan	0	Rendah
31	Rancamaya	0	Rendah
32	Ranggamekar	2	Tinggi
33	Babakan	0	Rendah
34	Babakan pasar	0	Rendah
35	Cibogor	0	Rendah
36	Ciwaringin	0	Rendah
37	Gudang	0	Rendah
38	Kebon Kelapa	0	Rendah
39	Pabaton	0	Rendah
40	Paledang	0	Rendah
41	Panaragan	0	Rendah
42	Sempur	0	Rendah
43	Tegallega	0	Rendah
44	Baranangsiang	1	Sedang
45	Katulampa	1	Sedang
46	Sindangrasa	0	Rendah
47	Sindangsari	0	Rendah
48	Sukasari	0	Rendah
49	Tajur	0	Rendah
50	Bantar Jati	1	Sedang
51	Cibuluh	1	Sedang

No	Kelurahan	*Cluster*	Information
52	Ciluar	2	Tinggi
53	Cimahpar	0	Rendah
54	Ciparigi	1	Sedang
55	Kedung Halang	2	Tinggi
56	Tanah Baru	1	Sedang
57	Tegal Gundil	1	Sedang
58	Cibadak	1	Sedang
59	Kasiyumanis	2	Tinggi
60	Kebon Pedes	1	Sedang
61	Kedung Badak	1	Sedang
62	Kedung Jaya	0	Rendah
63	Kedung Waringin	1	Sedang
64	Kencana	0	Rendah
65	Mekarwangi	0	Rendah
66	Sukadamai	0	Rendah
67	Sukaresmi	0	Rendah
68	Tanah sereal	0	Rendah

The results obtained for active IKM data for the city of Bogor in 2021 – 2022 with the application of the k-means clustering algorithm yield 3 clusters, with each cluster 0 criterion being the distribution of low IKM with a total of 45 sub-districts. Cluster 1 is the distribution of medium IKM with the number of sub-districts is 14, and cluster 2 is the distribution of high IKM with the number of sub-districts totaling 9. The results of grouping based on sub-districts in the city of Bogor can be seen in Table 5.

Table 5. Clustering Results

No	Kelurahan	*Cluster*	Information
1	Balumbang Jaya, Bubulak, Cilendek Barat, Cilendek Timur, Curug, Curug Mekar, Gunung Batu, Loji, Margajaya, Menteng, Pasir Jaya, Pasir Mulya, Semplak, Situ Gede, Bondongan, Genteng, Harjasari, Kertamaya, Lawang Gintung, Muarasari, Pakuan, Pamoyanan, Rancamaya, Babakan, Babakan pasar, Cibogor, Ciwaringin, Gudang, Kebon Kelapa, Pabaton, Paledang, Panaragan, Sempur, Tegallega, Sindangrasa, Sindangsari, Sukasari, Tajur, Cimahpar, Kedung Jaya, Kencana, Mekarwangi, Sukadamai, Sukaresmi, Tanah Sereal.	*Cluster 0*	The distribution of IKM is low
2	Cibuluh, Ciparigi, Tanah Baru, Tegal Gundil, Cibadak, Kebon Pedes, Kedung Badak, Kedung Waringin	Cluster 1	The distribution of IKM is medium
3	Batutulis, Bojong Kerta, Cikaret, Cipaku, Mulyaharja, Rangga Mekar, Ciluar, Kedung Halang, Kayu manis	Cluster 2	The distribution of IKM is high

3.4.1. Evaluation

The evaluation carried out in this study will produce interesting data information from the data that has been processed. The evaluation was carried out using scatter plot data visualization and cluster value testing using the silhouette coefficient method. The results of visualization and data testing in the study are shown in Figure 6.

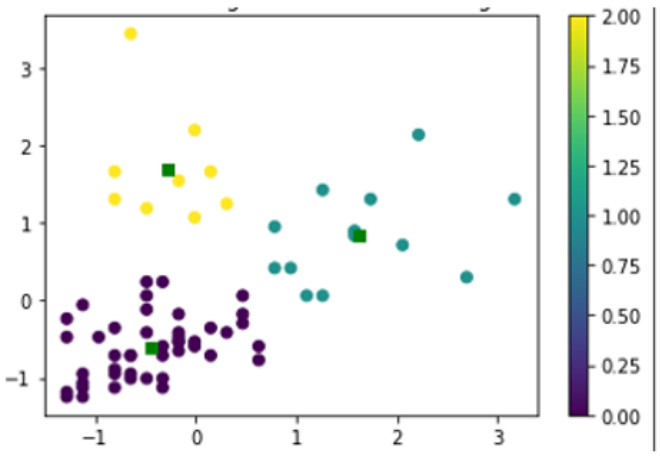

Figure 6. Visualization results using scatter plotss

The visualization above shows the distribution of active IKM data for the city of Bogor based on sub-districts from 2021 – 2022. The distribution of these points is visualized using a scatter plot. From the visualization above, there are 3 (three) colored dots, namely:
a. The purple color shows the distribution of data in cluster 0 (low) with a total of 45 points.
b. The blue color shows the distribution of data in cluster 1 (medium) with a total of 14 point distributions.
c. The yellow color shows the distribution of data in cluster 2 (high) with a total distribution of 9 points

This stage is also visualized using a pie chart, to find out the percentage of the number of groupings based on existing clusters out of 68. The percentage of clusters is shown in Figure 7.

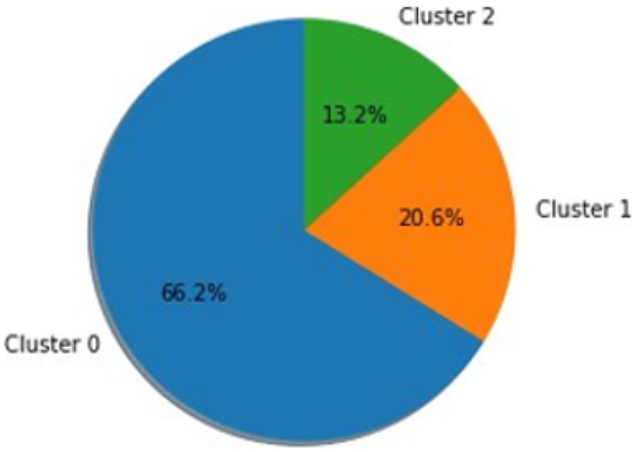

Figure 7. The percentage of clusters

The visualization results show that cluster 0 has a percentage of 66.2%, cluster 1 has a percentage of 20.6%, and cluster 2 has a percentage of 13.2%. So, in this visualization, 45 sub-districts have the most distribution of IKM, namely cluster 0 (low IKM distribution). Furthermore, testing the quality of the clusters that have been processed using the silhouette coefficient method is carried out. The results of cluster testing with python tools shown in Figure 8.

The evaluation results above show an index value of 0.56, which means entering into clustering criteria with good structure

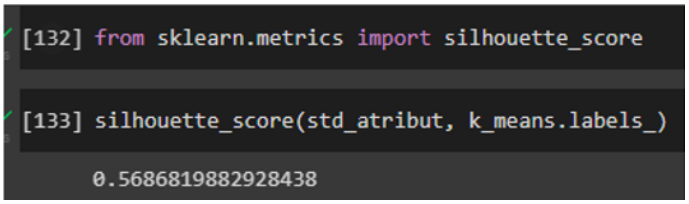

Figure 8. Test results using the silhouette coefficient method

3.5. Conclusion

Analysis of the distribution of active small and medium industries (IKM) in the city of Bogor from 2021 – 2022 using the k-means clustering method can result in grouping the distribution of SMI data based on urban villages in Bogor city with a total of 68 urban villages. The data processing process uses the python programming language and Google assistance collaboratory. Thus, the cluster results obtained from the application of this method were 3 clusters, namely cluster 0 included in the distribution of low IKM, cluster 1 included in the distribution of medium IKM, and cluster 2 included in the distribution of high IKM.

The results of grouping the distribution of active IKM data in the city of Bogor based on kelurahan, namely in cluster 0 with a low level of grouping, which means that the sub-districts that are included in cluster 0 have a small number of IKM, so agencies or the government need to improve in building and developing IKM in kelurahan - the kelurahan, cluster 0 consists of 45 kelurahan. Cluster 1 with a moderate level of grouping, which means that the sub-districts included in cluster 1 have a moderate number of IKM. So it is necessary to do development to increase IKM in the sub-districts that are included in the cluster, cluster 1 consists of 14 sub-districts. Cluster 2 with a high level of grouping, which means that the sub-districts included in cluster 2 have a large number of IKM. Kelurahan that have a large number of IKMs can contribute to boosting economic growth in the city of Bogor, cluster 2 consists of 9 kelurahan. From these results it is hoped that the government or related agencies can pay attention to the sub-district to increase the number of IKM in the sub-district and can provide related strategies in building and developing businesses in the sub-district.

The evaluation carried out in this study used the silhouette coefficient method for testing cluster quality, from the data used it produced a cluster value of 0.56 which means that it is included in the clustering criteria with a good structure.

References

[1] Sudrajat, W., Cholid, L, Petrus J. (2022). Penerapan Algoritma *K-Means* Clustering untuk Pengelompokkan UMKM menggunakan Rapidminer,*JUPITER, 27-36.*

[2] Dewi, S.P., Nurwati, Rahayu, E. (2022). Penerapan Data Mining untuk Prediksi Penjualan Produk terlarismenggunakan Metode *K-Nearest Neighbor. Technology and science, 639-648.*

[3] DKarsito, Sari, W. M. (2018). Prediksi Potensi Penjualan Produk Delifrance Dengan Metode *Naive Bayes* Di PT. PANGAN LESTARI.*Teknologi Pelita Bangsa, 67 - 78.*

[4] Nainggolan, R., Perangin-angin, R., Simarmata, E., Tarigan, F. A. (2019). *Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method. Physics.*

[5] Febrianti, A. F., Cabral, A. H., Anuraga, G. (2018). K-means Clustering Dengan Metode Elbow Untuk Pengelompokkan Kabupaten Dan Kota Di Jawa Timur Berdasarkan Indikator Kemiskinan. 863 - 870.

[6] Ridlo, M., Defiyanti, S., Primajaya, A. (2017). Implementasi Algoritme K-means Untuk Pemetaan Produktivitas Panen Padi Di Kabupaten Karawang. 426 - 432.

[7] Rahmah, S. A. (2020). Klasterisasi Pola Penjualan Pestisida Menggunakan Metode *K-means Clustering* (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja).*Information Technology Research, 1 - 5.*

[8] Utomo, W. (2021) *The comparison of and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia. 13*(1), 31–35.

[9] Heraldi, H. Y., Aprilia, N. C., Pratiwi, H. (2019). Analisis Cluster Intensitas Kebencanaan di Indonesia Menggunakan Metode *K-means. Indonesian Journal of Applied Statistics*, 137 - 144.

[10] Darnita, Y., Toyib, R., Kurniawan, Y. (2020). Penerapan Metode *K-means Clustering* Pada Aplikasi Android Pada Tanaman Obat Herbal. *Pseudocode*, 105 - 114.

[11] Aziz, F. N., Setiawan, B. D., Arwani, I. (2018). Implementasi Algoritma *K-means* untuk Klasterisasi Kinerja Akademik Mahasiswa. *Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2243 - 2251.