Jurnal Matematika MANTIK
Vol. 8, No. 2, December 2022, pp. 113-123

ISSN: 2527-3159 (*print*) 2527-3167 (*online*)

Simulation Of The K-Means Clustering Algorithm With The Elbow Method in Making Clusters Of Provincial Poverty Levels in Indonesia

Joko Riyono[1], Christina Eni Pujiastuti[1]

[1]Universitas Trisakti, Jakarta, Indonesia

Article history:
Received Sep 1, 2022
Revised Nov 30, 2022
Accepted Dec 31, 2022

Kata Kunci:
defisit anggaran, clustering, cluster optimal, kedalaman kemiskinan, keparahan kemiskinan

Abstrak. Salah satu cara agar progam dan bantuan pemerintah ke tiap provinsi bisa tepat sasaran adalah dengan membuat model pengelompokan atau clustering provinsi di Indonesia didasarkan pada tingkat kemiskinan. Algoritma K Means merupakan salah satu metode clustering di Data Mining untuk membagi n pengamatan menjadi k kelompok sedemikian hingga tiap pengamatan berada dalam kelompok dengan rata-rata terdekat. Dalam penelitian ini akan dibuat clustering tingkat kemiskinan Provinsi di Indonesia didasarkan pada tiga indikator tingkat kemiskinan yaitu Prosentase Penduduk Miskin (P0), Kedalaman Kemiskinan (P1) dan Keparahan Kemiskinan (P2) dengan Algoritma K-Means menggunakan Metode Elbow berbantukan Progam Phyton. Diperoleh hasil 5 Cluster optimal tingkat kemiskinan Provinsi di Indonesia.

Keywords:
Budget deficit, Clustering, Optimal cluster, Poverty depth, Poverty severity

Abstract. One way to ensure that government programs and assistance for each province are right on target is to create a model of grouping or clustering provinces in Indonesia based on poverty levels. Algorithm K Means is one of the clustering methods in Data Mining to divide n observations into k groups so that each observation is in the group with the closest mean. In this study, provincial poverty level clustering in Indonesia will be made based on three poverty level indicators, namely the Percentage of Poor Population (P0), Poverty Depth (P1), and Poverty Severity (P2) with the K-Means Algorithm using the Elbow Method assisted by the Python Program. The results obtained are 5 optimal clusters of provincial poverty rates in Indonesia.

How to cite:
J. Riyono and C. E. Pujiastuti, "Simulation of the K-Means Clustering Algorithm with The Elbow Method in Making Clusters of Provincial Poverty Levels in Indonesia", *J. Mat. Mantik*, vol. 8, no. 2, pp. 113-123, December 2022

CONTACT:
Joko Riyono jokoriyono@trisakti.ac.id Faculty of Industrial Technology, Universitas Trisakti,Indonesia

The article can be accessed here. *https://doi.org/10.15642/mantik.2022.8.2.113-123*

1. Introduction

The Covid-19 pandemic that occurred in Indonesia is part of a global pandemic which has a major impact on the country's economy. This pandemic has resulted in many companies going out of business, so many companies have laid off their employees. The large number of companies that have terminated employment has made unemployment more common in society, resulting in a decrease in people's purchasing power for goods and services, all of which have had an effect on increasing the level of poverty in society and slowing economic growth. According to Indonesian economic data from the Central Bureau of Statistics, the Indonesian economy will experience a decline of up to 2.41% in 2021.

The government has implemented various strategies in an effort to help the people's economic growth. The wide area coverage and the large number of people with different economic problems require different strategies to overcome them. There are many factors that serve as reference material for the actions that should be taken so that the assistance provided can be appropriate as needed. One way to ensure that government programs and assistance for each province are right on target is to create a model of grouping or clustering provinces in Indonesia based on poverty levels. The results of poverty mapping in the form of clustering are expected to provide benefits for the government in making policies and programs to overcome poverty problems and allocate budgets effectively and have a positive impact on poverty alleviation. In this study, provincial poverty level clustering in Indonesia will be made based on three poverty level indicators, namely the Percentage of Poor Population (P0), Poverty Depth (P1) and Poverty Severity (P2) with the K-Means Algorithm using the Elbow Method assisted by the Python Program.

Clustering analysis is a multivariate technique to group similar observations into a number of clusters based on the observed values of several variables for each individual. The purpose of the clustering process is to minimize the occurrence of the objective function that is set in the clustering process, which is generally used to minimize variations within a cluster and maximize variations between clusters[1]. One of the most frequently used algorithms in statistics and machine learning is K-means Clustering. K-Mean clustering is one of the unsupervised learning algorithms which is included in the non-hierarchical cluster analysis which is used to group data based on variables or features. The purpose of K-Means Clustering, like other Cluster methods, is to obtain clusters of data by maximizing the similarity of characteristics within the cluster and maximizing the differences between clusters. The K-means clustering algorithm groups data based on the distance between the data and the cluster centroid point obtained through an iterative process. The performance of the K-means clustering algorithm depends on the highly efficient cluster it forms. Determining the optimal number of clusters is an important step in creating Clusters with K-Means Clustering. There are several different ways to find the optimal number of clusters, one of the most popular methods for finding the number of clusters is the Elbow method.

The Elbow method is a method used to generate information in determining the best number of clusters by looking at the percentage of the comparison between the number of clusters that will form an angle at a point. This method uses the Within Cluster Sum of Squares (WCSS) value concept, which defines the total variation within a cluster. This method provides ideas or ideas by selecting the cluster value and then adding the cluster value to be used as a data model in determining the best cluster. And besides that the percentage of the resulting calculation becomes a comparison between the number of clusters added. The results of different percentages of each cluster value can be shown by using a graph as a source of information. Clustering is one method of data mining, data mining can be defined as the extraction of useful information or drawing patterns of knowledge from data stored in large quantities [2]. Data mining analysis runs on the best data and techniques to get the most feasible conclusions [3]. Data mining also has several names such as Knowledge Discovery in Database (KDD), knowledge extraction (Knowledge Extraction), business intelligence (Business Intelligence), and others.

One of the clustering algorithms is the K-Means clustering algorithm. K-Means clustering is a data analysis method or data mining method that performs a modeling process without supervision where k is a constant for the number of clusters and Means is the average value of a data group defined as a cluster [4].K-Means Clustering is one of the no hierarchical cluster analysis methods that group objects based on their characteristics so that they are in the same cluster and objects that have different characteristics are grouped together with similar objects in other clusters. The K-Means algorithm process can be described in a flowchart diagram or flow chart as follows [5]:

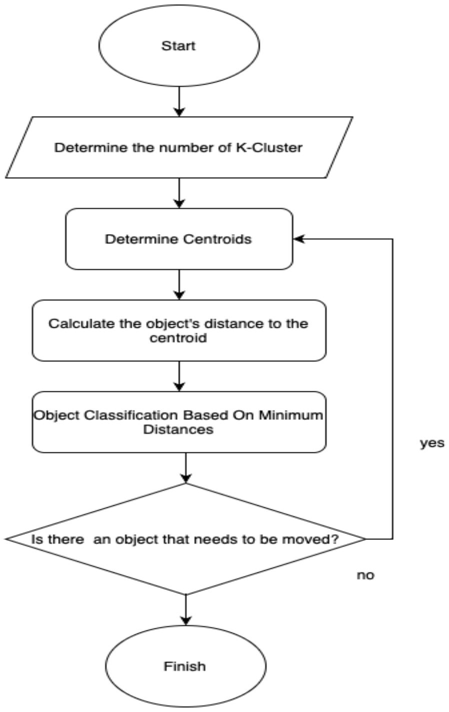

Figure 1. K-Means Algorithm Flowchart

Based on Figure 1, the initial stage of applying the K-Means Clustering algorithm is the selection of the value of k randomly based on the number of clusters needed, determining the centroid or center point of each cluster, calculating the distance of the object matrix to the center point or centroid, grouping each object into clusters based on distance. minimum or shortest, and iterate over each process starting from determining the centroid based on the final result of the temporary cluster. If there is no change in the group members of each cluster, the iteration can be stopped and the final results of clustering will be seen using the K-Means method.

An explanation of the stages of K-Means Clustering can be formulated as follows [6]:

a. Determine the value of k randomly
Determination of the value of k randomly is done based on the needs or desires of researchers. The value of k is used as the number of clusters to be formed. Determination of the value of k can be done through several considerations, both conceptual and theoretical [7].

b. Determine the center point or k centroid
Determination of the center point or initial centroid can be done randomly from the available objects. The number of central points determined must be based on the

number of k values because the center point in question is the center point of each cluster. In the iteration process, the determination of the i-th cluster centroid can be done by the formula [8]:

$$v=\frac{\Sigma_{i-1}^{n} x_i}{n} ;=1,2,3, \ldots n \ldots\ldots(1)$$

Information:
v : centroid on cluster
x_i : i-th object
n : the number of objects/ number of objects that are members of the cluster.

c. Calculate the distance of each object to each centroid
Calculation of the distance between each object and the centroid can be done by calculating the Euclidian Distance with the following formula:

$$(x,y)=\|x-y\|=\sqrt{\sum_{i-1}^{n}(x_i-y_i)^2}, i=1,2,3,\ldots n \ldots\ldots(2)$$

Information:
x_i : ith x object
y_i : y-th power
n : number of objects

d. Group each object into clusters
The grouping of each object into clusters is based on the minimum distance or the shortest distance that each object has to each center point or centroid. In the iteration process, grouping objects into each cluster can be done with Hard K-Means or Fuzzy C-Means [9].
e. Doing iteration
The iteration process is carried out from determining the centroid based on the final results of the temporary cluster and the next process is carried out as previously done. If in the new iteration there is no change in the cluster group, then the iteration can be stopped and the final results of the cluster can be seen. If the results are the same then the K-Means cluster analysis algorithm has converged, but if it is different, then it has not converged so it is necessary to do the next iteration [10]

2. Method

The method used in this study is a quantitative method using literacy studies, where data is collected using measuring instruments, then analyzed statistically and quantitatively. In accordance with the intent and purpose of the study, the data taken as a population is data on the Percentage of Poor Population, Poverty Depth and Poverty Severity in 34 provinces in Indonesia while the data taken as a sample is data on Percentage of Poor Population, Poverty Depth and Poverty Severity in 34 provinces in Semester I of 2021, Semester II of 2021 and Semester I of 2022. Data was taken through a literature study on the official website of the Central Statistics Agency with the website address https://bps.go.en/ [8]. Through this sample data will be analyzed to obtain clusters of population poverty levels in this case the provinces in Indonesia using the K-Means Clustering algorithm based on the Elbow method with Python software.

The selection of quantitative methods is based on existing research results, including: Quantitative methods emphasize objective measurements and statistical, mathematical or numerical analysis of the data obtained [11]. Research with quantitative methods has several advantages, including [12]:
a. Based on accurate data and measurements
b. Supported by advanced statistical techniques and software
c. Predictability validity that can be extended into the future

d. Can control at any time the validity of the relationship between the dependent and independent variables

	PROVINCES	A	B	C	D	E	F	G	H	I	INDICATOR P(0)	INDICATOR P(1)	INDICATOR P(2)
0	ACEH	15.33	15.53	14.64	2.86	2.95	2.49	0.75	0.81	0.61	15.166667	2.766667	0.723333
1	NORTH SUMATERA	9.01	8.49	8.42	1.52	1.45	1.36	0.38	0.38	0.34	8.640000	1.443333	0.366667
2	WEST SUMATERA	6.63	6.04	5.92	1.04	0.96	0.80	0.24	0.23	0.16	6.196667	0.933333	0.210000
3	RIAU	7.12	7.00	6.78	1.06	1.09	1.10	0.24	0.28	0.27	6.966667	1.083333	0.263333
4	JAMBI	8.09	7.67	7.62	1.29	1.09	1.17	0.30	0.23	0.26	7.793333	1.183333	0.263333
5	SOUTH SUMATERA	12.84	12.79	11.90	2.26	2.33	1.96	0.54	0.64	0.45	12.510000	2.183333	0.543333
6	BENGKULU	15.22	14.43	14.62	2.57	2.45	2.43	0.62	0.55	0.58	14.756667	2.483333	0.583333
7	LAMPUNG	12.62	11.67	11.57	1.88	1.85	1.82	0.42	0.38	0.41	11.953333	1.850000	0.403333
8	BANGKA BELITUNG ISLANDS	4.90	4.67	4.45	0.61	0.56	0.60	0.11	0.11	0.13	4.673333	0.590000	0.116667
9	RIAU ISLANDS	6.12	5.75	6.24	1.07	0.95	1.05	0.28	0.22	0.25	6.036667	1.023333	0.250000
10	JAKARTA	4.72	4.67	4.69	0.64	0.75	0.77	0.14	0.18	0.19	4.693333	0.720000	0.170000
11	WEST JAVA	8.40	7.97	8.06	1.47	1.29	1.32	0.38	0.31	0.33	8.143333	1.360000	0.340000
12	CENTRAL JAVA	11.79	11.25	10.93	1.91	1.94	1.77	0.45	0.46	0.42	11.323333	1.873333	0.443333
13	SPECIAL REGION OF YOGYAKARTA	12.80	11.91	11.34	2.42	2.06	2.01	0.65	0.53	0.51	12.016667	2.163333	0.563333
14	EAST JAVA	11.40	10.59	10.38	1.84	1.58	1.62	0.43	0.33	0.38	10.790000	1.680000	0.380000
15	BANTEN	6.66	6.50	6.16	1.09	1.20	1.03	0.26	0.34	0.27	6.440000	1.106667	0.290000
16	BALI	4.53	4.72	4.57	0.68	0.76	0.62	0.15	0.17	0.13	4.606667	0.686667	0.150000
17	WEST NUSA TENGGARA	14.14	13.83	13.68	2.24	2.50	2.49	0.49	0.63	0.67	13.883333	2.410000	0.596667
18	EAST NUSA TENGGARA	20.99	20.44	20.05	3.96	4.79	3.63	1.05	1.44	0.93	20.493333	4.126667	1.140000
19	WEST KALIMANTAN	7.15	6.84	6.73	1.03	1.02	1.04	0.23	0.24	0.24	6.906667	1.030000	0.236667
20	CENTRAL KALIMANTAN	5.16	5.16	5.28	0.71	0.75	0.91	0.15	0.17	0.27	5.200000	0.790000	0.196667
21	SOUTH KALIMANTAN	4.83	4.56	4.49	0.67	0.55	0.63	0.13	0.10	0.13	4.626667	0.616667	0.120000
22	EAST KALIMANTAN	6.54	6.27	6.31	1.22	1.04	0.99	0.34	0.23	0.23	6.373333	1.083333	0.266667
23	NORTH KALIMANTAN	7.36	6.83	6.77	0.87	0.89	0.89	0.18	0.17	0.19	6.986667	0.883333	0.180000
24	NORTH SULAWESI	7.77	7.36	7.28	1.24	1.04	1.15	0.28	0.22	0.26	7.470000	1.143333	0.253333
25	CENTRAL SULAWESI	13.00	12.18	12.33	2.43	2.24	2.41	0.64	0.62	0.68	12.503333	2.360000	0.646667
26	SOUTH SULAWESI	8.78	8.53	8.63	1.49	1.40	1.36	0.37	0.34	0.32	8.646667	1.416667	0.343333
27	SOUTHEAST SULAWESI	11.66	11.74	11.17	2.16	2.40	1.82	0.56	0.69	0.42	11.523333	2.126667	0.556667
28	GORONTALO	15.61	15.41	15.42	2.87	2.92	3.04	0.72	0.77	0.85	15.480000	2.943333	0.780000
29	WEST SULAWESI	11.29	11.85	11.75	1.76	1.90	2.21	0.44	0.50	0.58	11.630000	1.956667	0.506667
30	MALUKU	17.87	16.30	15.97	3.58	3.49	2.90	1.05	1.06	0.80	16.713333	3.323333	0.970000
31	NORTH MALUKU	6.89	6.38	6.23	0.97	0.94	0.91	0.21	0.20	0.20	6.500000	0.940000	0.203333
32	WEST PAPUA	21.84	21.82	21.33	5.49	5.84	4.82	1.96	2.18	1.60	21.663333	5.383333	1.913333
33	PAPUA	26.86	27.38	26.56	5.60	6.31	6.16	1.69	2.05	2.10	26.933333	6.023333	1.946667

Figure 2. Poverty Percentage Data, Poverty Depth Index Data, and Poverty Severity Index Data in 34 provinces in Indonesia in the first semester of 2021, second semester of 2021 and first semester of 2022

Caption:
Column A: the percentage of poor population (P0) semester I 2021
Column B: the percentage of poor population (P0) semester II 2021
Column C: the value of the percentage of poor people (P0) in the first semester of 2022
Column D: Poverty Depth score (P1) semester I 2021
Column E: Poverty Depth score (P1) in semester II 2021
Column F: Poverty Depth score (P1) semester I 2022
Column G: Poverty Severity score (P2) semester I 2021
Column H: Poverty Severity score (P2) semester II 2021
Column I: Poverty Severity score (P2) semester I 2022
Indicator column P(0) : Average value of the percentage of poor population (P0) 2021-2022
Indicator column P(1) : Average value of Poverty Depth (P1) 2021-2022
Indicator column P(2) : Average score of Poverty Severity (P2) 2021-2022

Standardization is one of the processing processes by equating the units of each attribute. In cases where the data has a much different range of values it can cause the calculation to be ineffective [14].So standardization is done so that the data has the same range of values so it is hoped that data with large or small values will not affect the final analysis results. One way to standardize data is to use Z-Score, basically standardizing using Z-Score is to change the original data value into Z form (normally distributed data) with the formula [15]:

$$z = \frac{x_{i-\bar{x}}}{S} \text{ with } \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \text{ and } S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{...(3)}$$

3. Results and Discussion

The research process with the K-Means Clustering algorithm used in this study is to group 34 provinces in Indonesia based on poverty level indicators from semester I 2021, semester II 2021 and semester I 2022 with the help of Python software. The following are the stages in the research process:

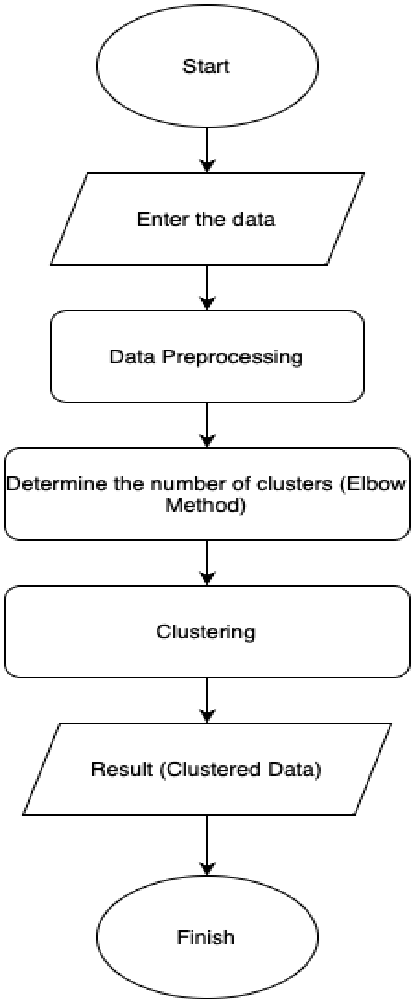

Figure 3. Research Flowchart

a. Input the data set Percentage of Poor Population (P0), Poverty Depth Index P(1), and Poverty Severity Index P(2) found in 34 provinces in Indonesia in semester I 2021, semester II 2021 and semester I 2022 using Microsoft Excel.
b. Pre-processing data with Z-Score Normalization method

	p(0)	p(1)	p(2)
0	0.889672	0.704713	0.525219
1	-0.348666	-0.338980	-0.304458
2	-0.812253	-0.741209	-0.668896
3	-0.666157	-0.622906	-0.544832
4	-0.509309	-0.544038	-0.544832
5	0.385609	0.244647	0.106503
6	0.811881	0.481252	0.199551
7	0.279990	-0.018248	-0.219164
8	-1.101283	-1.011991	-0.886007
9	-0.842610	-0.670227	-0.575848
10	-1.097488	-0.909462	-0.761944
11	-0.442901	-0.404703	-0.366490
12	0.160456	0.000155	-0.126116
13	0.292006	0.228873	0.153027
14	0.059264	-0.152324	-0.273442
15	-0.766084	-0.604504	-0.482800
16	-1.113932	-0.935751	-0.808467
17	0.646179	0.423415	0.230567
18	1.900328	1.777324	1.494467
19	-0.677541	-0.664969	-0.606864
20	-1.001355	-0.854254	-0.699912
21	-1.110137	-0.990959	-0.878253
22	-0.778733	-0.622906	-0.537078
23	-0.662362	-0.780643	-0.738682
24	-0.570656	-0.575585	-0.568094
25	0.384344	0.383981	0.346877
26	-0.347401	-0.360011	-0.358736
27	0.198403	0.199955	0.137519
28	0.949122	0.844047	0.657037
29	0.218642	0.065878	0.021209
30	1.183129	1.143747	1.099014
31	-0.754700	-0.735951	-0.684404
32	2.122318	2.768437	3.293393
33	3.122223	3.273195	3.370933

Figure 4. Results of pre-processing data

c. Determine the value of k clusters or the number of clusters through the Elbow method. Elbow method is a heuristic used in determining the number of clusters in a data set by plotting the variation described as a function of the number of clusters and selecting the angle of the curve as the number of clusters to be used.

	Clusters	WSS
0	1	102.000000
1	2	35.305426
2	3	11.463562
3	4	6.091538
4	5	4.380479
5	6	2.923555
6	7	2.287246
7	8	1.656945
8	9	1.098996
9	10	0.750493
10	11	0.489641
11	12	0.393661
12	13	0.316045
13	14	0.268826
14	15	0.208264
15	16	0.181340
16	17	0.150592
17	18	0.128350
18	19	0.110715
19	20	0.087626
20	21	0.075049
21	22	0.061230
22	23	0.046301
23	24	0.034658
24	25	0.025815
25	26	0.017504
26	27	0.012584
27	28	0.009711
28	29	0.007061
29	30	0.005271
30	31	0.003548
31	32	0.001853
32	33	0.000290
33	34	0.000000

Figure 5. WSS data value

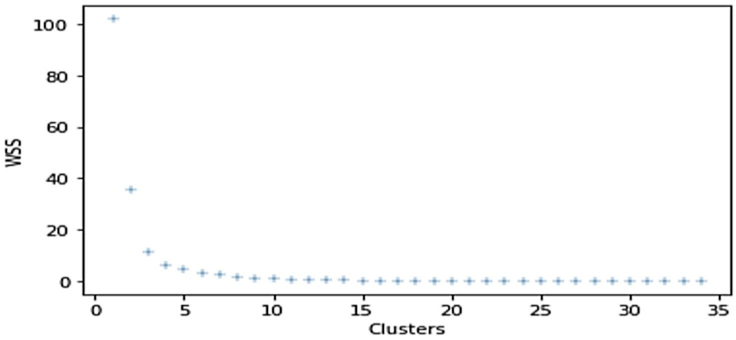

Figure 6. WSS vs cluster

from Figure 5 and 6 with the elbow method, the elbow point is found in cluster 5, therefore it is concluded that the number of clusters is 5.

d. Processing the results of pre-processing data in step 2 using the K-Means algorithm in Python software by inputting the value of k=5 or the number of clusters obtained from step 3.

	PROVINCES	A	B	C	D	E	F	G	H	I	INDICATOR P(0)	INDICATOR P(1)	INDICATOR P(2)	Clusters
0	ACEH	15.33	15.53	14.64	2.86	2.95	2.49	0.75	0.81	0.61	15.166667	2.766667	0.723333	4
1	NORTH SUMATERA	9.01	8.49	8.42	1.52	1.45	1.36	0.38	0.38	0.34	8.640000	1.443333	0.366667	0
2	WEST SUMATERA	6.63	6.04	5.92	1.04	0.96	0.80	0.24	0.23	0.16	6.196667	0.933333	0.210000	0
3	RIAU	7.12	7.00	6.78	1.06	1.09	1.10	0.24	0.28	0.27	6.966667	1.083333	0.263333	0
4	JAMBI	8.09	7.67	7.62	1.29	1.09	1.17	0.30	0.23	0.26	7.793333	1.183333	0.263333	0
5	SOUTH SUMATERA	12.84	12.79	11.90	2.26	2.33	1.96	0.54	0.64	0.45	12.510000	2.183333	0.543333	2
6	BENGKULU	15.22	14.43	14.62	2.57	2.45	2.43	0.62	0.55	0.58	14.756667	2.483333	0.583333	4
7	LAMPUNG	12.62	11.67	11.57	1.88	1.85	1.82	0.42	0.38	0.41	11.953333	1.850000	0.403333	2
8	BANGKA BELITUNG ISLANDS	4.90	4.67	4.45	0.61	0.56	0.60	0.11	0.11	0.13	4.673333	0.590000	0.116667	0
9	RIAU ISLANDS	6.12	5.75	6.24	1.07	0.95	1.05	0.28	0.22	0.25	6.036667	1.023333	0.250000	0
10	JAKARTA	4.72	4.67	4.69	0.64	0.75	0.77	0.14	0.18	0.19	4.693333	0.720000	0.170000	0
11	WEST JAVA	8.40	7.97	8.06	1.47	1.29	1.32	0.38	0.31	0.33	8.143333	1.360000	0.340000	0
12	CENTRAL JAVA	11.79	11.25	10.93	1.91	1.94	1.77	0.45	0.46	0.42	11.323333	1.873333	0.443333	2
13	SPECIAL REGION OF YOGYAKARTA	12.80	11.91	11.34	2.42	2.06	2.01	0.65	0.53	0.51	12.016667	2.163333	0.563333	2
14	EAST JAVA	11.40	10.59	10.38	1.84	1.58	1.62	0.43	0.33	0.38	10.790000	1.680000	0.380000	2
15	BANTEN	6.66	6.50	6.16	1.09	1.20	1.03	0.26	0.34	0.27	6.440000	1.106667	0.290000	0
16	BALI	4.53	4.72	4.57	0.68	0.76	0.62	0.15	0.17	0.13	4.606667	0.686667	0.150000	0
17	WEST NUSA TENGGARA	14.14	13.83	13.68	2.24	2.50	2.49	0.49	0.63	0.67	13.883333	2.410000	0.596667	4
18	EAST NUSA TENGGARA	20.99	20.44	20.05	3.96	4.79	3.63	1.05	1.44	0.93	20.493333	4.126667	1.140000	1
19	WEST KALIMANTAN	7.15	6.84	6.73	1.03	1.02	1.04	0.23	0.24	0.24	6.906667	1.030000	0.236667	0
20	CENTRAL KALIMANTAN	5.16	5.16	5.28	0.71	0.75	0.91	0.15	0.17	0.27	5.200000	0.790000	0.196667	0
21	SOUTH KALIMANTAN	4.83	4.56	4.49	0.67	0.55	0.63	0.13	0.10	0.13	4.626667	0.616667	0.120000	0
22	EAST KALIMANTAN	6.54	6.27	6.31	1.22	1.04	0.99	0.34	0.23	0.23	6.373333	1.083333	0.266667	0
23	NORTH KALIMANTAN	7.36	6.83	6.77	0.87	0.89	0.89	0.18	0.17	0.19	6.986667	0.883333	0.180000	0
24	NORTH SULAWESI	7.77	7.36	7.28	1.24	1.04	1.15	0.28	0.22	0.26	7.470000	1.143333	0.253333	0
25	CENTRAL SULAWESI	13.00	12.18	12.33	2.43	2.24	2.41	0.64	0.62	0.68	12.503333	2.360000	0.646667	2
26	SOUTH SULAWESI	8.78	8.53	8.63	1.49	1.40	1.36	0.37	0.34	0.32	8.646667	1.416667	0.343333	0
27	SOUTHEAST SULAWESI	11.66	11.74	11.17	2.16	2.40	1.82	0.56	0.69	0.42	11.523333	2.126667	0.556667	2
28	GORONTALO	15.61	15.41	15.42	2.87	2.92	3.04	0.72	0.77	0.85	15.480000	2.943333	0.780000	4
29	WEST SULAWESI	11.29	11.85	11.75	1.76	1.90	2.21	0.44	0.50	0.58	11.630000	1.956667	0.506667	2
30	MALUKU	17.87	16.30	15.97	3.58	3.49	2.90	1.05	1.06	0.80	16.713333	3.323333	0.970000	1
31	NORTH MALUKU	6.89	6.38	6.23	0.97	0.94	0.91	0.21	0.20	0.20	6.500000	0.940000	0.203333	0
32	WEST PAPUA	21.84	21.82	21.33	5.49	5.84	4.82	1.96	2.18	1.60	21.663333	5.383333	1.913333	3
33	PAPUA	26.86	27.38	26.56	5.60	6.31	6.16	1.69	2.05	2.10	26.933333	6.023333	1.946667	3

Figure 7. Results of K-Means Clustering with python software

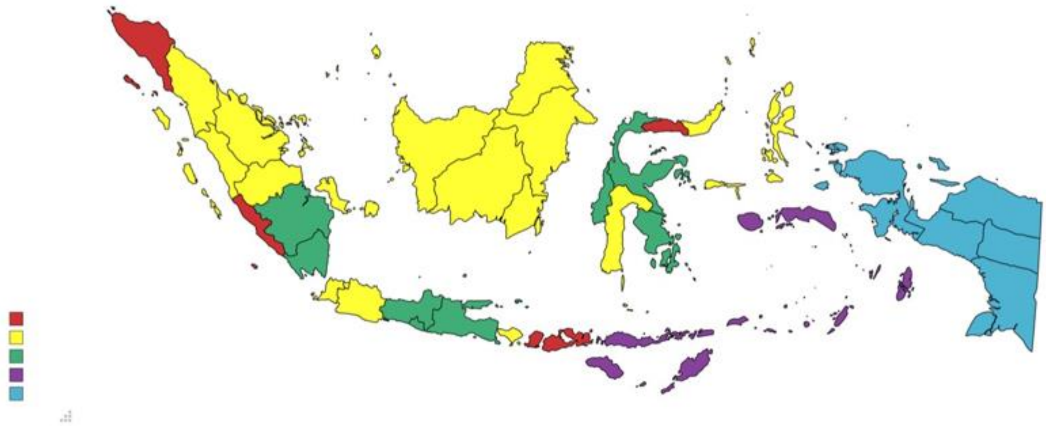

Figure 8. Image of the Clustering map of the poverty level of the Indonesian province with K means Clustering using the Python software

From Figures 7 and 8, the results show that there are 5 optimal clusters of provincial poverty levels in Indonesia obtained by the elbow method, the five clusters are: The first cluster contains the provinces of Papua and West Papua with an average poverty rate indicator of 10.644. The second cluster contains the provinces of Maluku and East Nusa Tenggara with an average poverty rate indicator of 7.79. the third cluster contains the provinces of Aceh, Bengkulu, West Nusa Tenggara, Gorontalo with an average poverty rate indicator of 4.5635. The fourth cluster contains the provinces of South Sumatra, Lampung, Central Java, Yogyakarta, East Java, Central Sulawesi, Southeast Sulawesi, West Sulawesi with an average poverty rate indicator of 4.77 and the fifth cluster contains the provinces of North Sumatra, West Sumatra, Riau, Jambi, Bangka Islands. Belitung, Riau Islands, Jakarta, West Java, Banten, Bali, West Kalimantan, South Kalimantan, Central Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, South Sulawesi, North Maluku with an average poverty rate indicator of 2.58. Based on the average population poverty rate, it can be seen that the provinces of Papua and West Papua are two provinces that need attention, even though in terms of natural resources the two provinces have very abundant wealth, this can be used as a topic for further research on why conditions like this can occur.

4. Conclusion

The conclusion that can be drawn from the grouping of provincial poverty levels in Indonesia using the Elbow method is that there are 5 poverty level clusters. Then using the K-Means Clustering algorithm, clusters with very high poverty rates are found in the provinces of Papua and West Papua. High poverty rates are found in Maluku and East Nusa Tenggara. Poverty levels are moderate in the provinces of Aceh, Bengkulu, West Nusa Tenggara, Gorontalo. Low poverty rates are found in South Sumatra, Lampung, Central Java, Yogyakarta, East Java, Central Sulawesi, Southeast Sulawesi, West Sulawesi and very low poverty rates are found in the provinces of North Sumatra, West Sumatra, Riau, Jambi, Bangka Islands. Belitung, Riau Islands, Jakarta, West Java, Banten, Bali, West Kalimantan, South Kalimantan, Central Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, South Sulawesi, North Maluku.

References

[1] M. Safar, "K-Means-Implementation, Problems and Related Methods"

[2] W.Priatna, “Data Mining Lecture Module - Ubharajaya Repository.” http://repository.ubharajaya.ac.id/6318/ (accessed Aug. 31, 2022).

[3] R. Fdillah, "Implementation of Data Mining in the Analysis of Fire Incidents in the City of Bandung Using the Association Rule Method - Elibrary Unikom." https://elibrary.unikom.ac.id/id/eprint/2686/ (accessed Aug. 31, 2022).

[4] Nasari, “Implementation of k-means clustering on new student admission data (case study: potential university utama) Online SEMNASTEKNOMEDIA.” https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/view/837 (accessed Aug. 31, 2022).

[5] S. Andayani, "Cluster Formation in Knowledge Discovery in Database with K-Means Algorithm - Lumbung Pustaka UNY." https://eprints.uny.ac.id/2354/ (accessed Aug. 31, 2022).

[6] A. Asroni, H. Fitri, and E. Prasetyo, "Implementation of the Clustering Method with the K-Means Algorithm in Grouping Prospective New Student Data at Yogyakarta Muhammadiyah University (Case Study: Faculty of Medicine and Health Sciences, and Faculty of Social and Political Sciences),” *Semesta Teknika*, vol. 21, no. 1, pp. 60–64, May 2018, doi: 10.18196/ST.211211

[7] M. Benri, H. Metisen, and S. Latipa, "Clustering analysis using the k-means method in product sales clustering at fadhila superiorities," *JURNAL MEDIA INFOTAMA*, vol. 11, no. 2, Sept. 2015, doi: 10.37676/JMI.V11I2.258.

[8] Y. Darmi, A. Setiawan, J. Bali, K. Kampung Bali, K. Teluk Segara, and K. Bengkulu, "Applying the K-Means clustering method in product sales clustering," *JURNAL MEDIA INFOTAMA*, vol. 12, no. 2, Dec. 2016, doi: 10.37676/JMI.V12I2.418.

[9] N. Shi, X. Liu, and Y. Guan, “Research on k-means clustering algorithm: An improved k-means clustering algorithm,” *3rd International Symposium on Intelligent Information Technology and Security Informatics, IITSI 2010*, pp. 63–67, 2010, doi: 10.1109/IITSI.2010.74.

[10] Neva Satyahadewi, “Characteristics classification with k-means cluster analysis method “Bimaster: Scientific Bulletin of Mathematics, Statistics and Its Applications.” https://jurnal.untan.ac.id/index.php/jbmstr/article/view/3033/2998 (accessed Aug. 31, 2022).

[11] R.V. Labaree, “Quantitative Methods - Organizing Your Social Sciences Research Paper - Research Guides at University of Southern California.” https://libguides.usc.edu/writingguide/quantitative (accessed Aug. 31, 2022).

[12] Z. Nabila, A. R. Isnain, P. Permata, and Z. Abidin, "Data mining analysis for clustering case of covid-19 in Lampung province with the K-Means algorithm," *Journal of Technology and Information Systems*, vol. 2, no. 2, pp. 100–108, Jul. 2021, doi: 10.33365/JTSI.V2I2.868.

[13] "Central Bureau of Statistics." ttps://www.bps.go.id/subject/23/kemiskinan-dan-ketimpangan.html#subjekViewTab3 (accessed Aug. 31, 2022).

[14] E. S. Y. Pandie, “Implementation of naive bayes data mining algorithms in cooperatives,” *J-ICON*, vol. 6, no. 1, pp. 15–20, 2018.

[15] T. Alfina, T. Alfina, B. Santosa, and A. R. Barakbah, "A Comparative Analysis of Hierarchical Clustering Methods, K-Means and the Combined Both in Cluster Data (Case Study: ITS Industrial Engineering Practice Work Problems)," *ITS Engineering Journal*, vol. 1, no. 1, pp. A521–A525, Sept. 2012, doi: 10.12962/j23373539.v1i1.1794.