BAREKENG: Journal of Mathematics and Its Applications
September 2023 Volume 17 Issue 3 Page 1663–1672
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss3pp1663-1672

# NON HIERARCHICAL K-MEANS ANALYSIS TO CLUSTERING PRIORITY DISTRIBUTION OF FUEL SUBSIDIES IN INDONESIA

**Ani Budi Astuti[1*], Abdi Negara Guci[2], Viky Iqbal Azizul Alim[3], Laila Nur Azizah[4], Meirida Karisma Putri[5], Wigbertus Ngabu[6]**

*[1,3,4,6]Department of Statistics, Faculty of Mathematics and Natural Sciences, Brawijaya University*
*[2,5]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Brawijaya University*
*Malang, 65144, Indonesia*

*Corresponding author's e-mail: *ani_budi@ub.ac.id*

## ABSTRACT

***Article History:***

*Received: 03rd May 2023*
*Revised: 9th August 2023*
*Accepted: 19th August 2023*

***Keywords:***

*Cluster analysis;*
*Elbow;*
*Inflation;*
*K-Means;*
*Oil subsidie;*
*Shilouette.*

*The growth rate of inflation in Indonesia continues to increase from day to day. The inflation rate in Indonesia reached 1.17% in September 2022 which is the highest inflation rate in the last seven years. One of the causes of high inflation is caused by the increasing demand for motor vehicle fuel. Therefore, there is a need for appropriate action from the government in determining related policies. K-Means multivariate cluster analysis is a non-hierarchical cluster method that is popularly used, one of which is used in Machine Learning algorithms, especially Unsupervised Learning. The purpose of this research is to clustering that are priority distribution of subsidies in Indonesia based on the characteristics formed. The data in this study consist of the percentage of poverty, the percentage of total transportation, the percentage of transportation use, and the percentage of area. Data were analyzed using multivariate cluster analysis with the K-Means method. Based on the research results, information was obtained that the data fulfilled a representative sample with value of KMO >50%. In addition, there are 4 optimal clusters which are the results of the calculation of the Elbow and Silhoutte methods, so 4 provincial clusters are formed with their respective characteristics. Cluster 1 is a province that is highly prioritized to receive fuel subsidies, Cluster 2 is a province that is not highly prioritized for fuel subsidies, Cluster 3 is a province that is prioritized to receive fuel subsidies, and Cluster 4 is a province that is not prioritized to receive fuel subsidies.*



***How to cite this article:***

A. B. Astuti, A. N. Guci, V. I. A. Alim, L. N. Azizah, M. K. Putri and W. Ngabu., "NON HIERARCHICAL K-MEANS ANALYSIS TO CLUSTERING PRIORITY DISTRIBUTION OF FUEL SUBSIDIES IN INDONESIA," *BAREKENG: J. Math. & App.*, vol. 17, iss. 3, pp. 1663-1672, September, 2023.


Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*
**Research Article · Open Access**

# 1. INTRODUCTION

Economic development is currently a global concern in assessing the success of a country's development. The development of economic development as a new scientific discipline brought about important changes in the level of population welfare. The level of population welfare is empirically correlated with the inflation rate in each country. Indonesia, one of the many developing countries, has the characteristics of inflation which is still relatively high and changes rapidly (volatile). Of course, this will lead to a decrease in the level of community prosperity. In addition, inflation also reduces productive investment, increases production costs, and causes economic instability [1]. Furthermore, in September 2022 there was inflation of 1.17% with a Consumer Price Index (CPI) of 112.87. The largest inflation rate occurred in September 2022, which was 1.17%. The inflation rate in Indonesia from January 2019 to December 2021 is still very volatile. However, starting in 2022 it shows a continuous increase in every month. The peak will be in September 2022 with a year-on-year inflation rate (September 2022 against September 2021) of 5.95%. This figure is the highest inflation rate in the last 7 years. This occurred due to price increases as indicated by increases in most of the expenditure group indices.

The Consumer Price Index (CPI) is an indicator that is often used to measure the inflation rate [2]. In simple terms, the Consumer Price Index (CPI) or also known as the Consumer Price Index (CPI) is a comparison between the prices of a package of commodities from a group of goods or services (market basket) consumed by households (households) in a certain period [3]. CPI inflation in Indonesia in the last 10 years has been more influenced by spikes in administered prices and volatile food inflation. Based on the Consumer Price Index (CPI) metadata published by Bank Indonesia in 2016, the housing, water, electricity, gas, and fuel group is the CPI group that has the highest market share percentage compared to other CPI groups. This group plays a very important role in the country's economy and has a major impact on inflation. Based on data from the Central Statistics Agency (BPS), the Consumer Price Index (CPI) for the transportation group in September 2022 was at the level of 119.64; increased by 8.88 percent compared to the previous month. This proves that 77% of the cause of inflation comes from the vehicle group [4]. The high inflation rate in the vehicle category was caused by soaring prices for fuel oil (BBM) in the oil and gas sector.

The crisis in fuel oil that occurred in Indonesia was marked by the difficulty for the public to obtain fuel and a significant increase in fuel prices [5]. The increase in world oil prices is considered to have an elusive impact on Indonesia's economic situation. High oil prices are certainly not good for the economy of any country, including Indonesia. In addition, the increase in fuel prices in Indonesia was also caused by a lack of supply (undersupply) in the country when compared to demand (demand) [6]. The amount of fuel subsidies in Indonesia which is a burden on the state budget is also considered to be the cause of the government increasing fuel prices in the country. Therefore, in August 2020, the Indonesian Political Indicator (IPI) surveyed the distribution of fuel subsidies by the government with the survey results showing that most Indonesians agree that fuel subsidies are not on target with 53.3% of respondents. The details are as much as 35.6% of respondents said they agreed and 17.7% of respondents said they strongly agreed. Of course, this is a fundamental thing that more fuel subsidies are enjoyed by people who are able and who can afford to buy motorized vehicles. Therefore, the government needs efforts and implementation of other policies to continue to control and even reduce the inflation rate due to increased transportation use in Indonesia, given the high inflation in September 2022 due to several factors, one of which is the distribution of fuel subsidies that are uneven and not on target.

Provision of subsidies from the government of course must be right on target and the government must pay attention to the needs of the community for the use of vehicles. Of course, every city has different needs which are influenced by various factors. Therefore, government policies should need to be designed with a scheme through comprehensive data analysis, both official and non-official data that can be accounted for so that appropriate and relevant policies can be obtained [7]. In applying a simple analysis to the problem of soaring inflation due to the high level of transportation groups in Indonesia, cluster analysis can be carried out. With an analysis that considers many variables without the response variable labels and predictor variables, where all variables have the same position. Cluster Multivariate Analysis is a statistical method used to group objects or individuals based on several different variables or features. The aim is to find patterns and structures in data that have multiple variables, evaluate the quality of the clustering model, and compare different groups to understand the correlation between variables [8].

Research related to multivariate analysis using K-Means clusters has been done before. Research conducted by Dini and Fauzan to classify provinces in Indonesia based on indicators of community welfare using K-Means cluster analysis [9] . The K-Means cluster analysis was chosen based on the variance value (0.101) which is smaller than the variance value in the average linkage cluster analysis (0.152). The results of the analysis show that the provinces in Indonesia are formed into three groups with different characteristics. Conducted a study entitled K-Means Cluster Analysis for Grouping Districts in South Sulawesi Province Based on Village Potential which discussed the grouping of sub-districts in South Sulawesi based on the potential in each village using multivariate analysis with K-Means analysis [10]. The conclusion of this study is the formation of three groups of sub-districts in South Sulawesi, namely Group 1 consisting of 107 sub-districts with high village potential, group 2 consisting of 16 sub-districts with moderate village potential, and Group 3 consisting of 184 sub-districts with low village potential. Research conducted discusses the grouping of trucks in the field of transportation services by using the K-Means, K-Medoids, and Davies Bouldin Index Algorithms as a cluster analysis method to produce validity values for cluster results [11]. The results obtained show that the K-Means algorithm cluster results are the most valid and relevant with a Davies Bouldin Index validity value of 0.67 and a concordance rate of 97%.

Based on various previous studies that have been conducted, it is proposed that this research is important and urgent to discuss provincial groupings based on priority distribution of vehicle subsidies in Indonesia using multivariate analysis with K-Means cluster analysis and cluster validity with the Elbow method and the Silhoutte method. The results of this study can be used to provide recommendations for the Government of Indonesia in determining policies for providing fuel subsidies that are right on target in various provinces in Indonesia based on the priority level of the grouping results, to reduce public complaints and unrest related to fuel subsidies provided by the Government of Indonesia as a form of real assistance to the people of Indonesia who need it.

## 2. RESEARCH METHODS

### 2.1 Data

The data used for this research is secondary data consisting of Poverty Percentage data ($X_1$) obtained from the Central Statistics Agency's website for recording on March 7, 2022, Transportation Amount Percentage data ($X_2$) obtained from the National Police Corps dashboard on recording October 26, 2022, Data on the Percentage of Use of Transportation ($X_3$) were obtained from the website of the Central Statistics Agency with records on October 7, 2022, and data on the Percentage of Areas ($X_4$) were obtained from the website of the Central Statistics Agency for 2021 records. This research period was carried out from March 2022 to October 2022. **Table 1** presents the categories of research variables.

**Table 1. Research Variable Categories**

| Variable | Measurement Scale |
|---|---|
| Poverty Percentage ($X_1$) | Ratio |
| Percentage of Total Transport ($X_2$) | Ratio |
| Percentage of Transportation Use ($X_3$) | Ratio |
| Territory Area ($X_4$) | Ratio |

### 2.2 Data Normalization

Data normalization is a form of transforming the original variable ($X$) to a standard number ($Z$) [12]. In addition, data normalization can be said as another technique of making changes to the scale of the data. In this study, normalization of data variable values is used to transform data into standard numbers. Mathematical formula for data normalization can be written as **Equation (1)** [13],

$$Z_i = \frac{X_i - \bar{X}}{s} \tag{1}$$

where:

$Z_i$ : the raw value of the observed data $X_i$

$X_i$ : observation data $X_i$

$\bar{X}$ : the average value of X observation data s : standard deviation value of X observation data

### 2.3 Test Assumptions

Two assumptions must be met in cluster analysis, namely a representative sample using the KMO test (Kaiser Meyer Olkin) [14].

**Representative Sample**

The KMO value (Kaiser-Meyer-Olkin) is a statistic used to measure the level of adequacy of data for factor analysis [15]. The values range from 0 to 1, where a value of 1 indicates that the data is very suitable for factor analysis, while a value of 0 indicates that the data is not suitable at all [16]. The sample is said to be appropriate if the KMO value ranges from 0.5 to 1.0 and vice versa if the KMO value is less than 0.5, then the sample is not representative of the population. The calculation of the KMO assumption test statistic (Kaiser Meyer Olkin) is defined in Equation (2) [17].

$$KMO = \frac{\left(\sum_{i=1}^{p}\sum_{j=1}^{p} r_{ij}^2\right)}{\sum_{i=1}^{p}\sum_{j=1}^{p} r_{ij}^2 + \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}^2} \tag{2}$$

where:

for $i \neq j, i = 1,2,\dots,p\,; j = 1,2,\dots,p$

$r_{ij}$ : simple correlation coefficient between variables

$a_{ij}$ : partial correlation coefficient between variables

### 2.4 Determining the Optimal K Value

a. Elbow method

The Elbow method is a technique used to determine the optimal number of clusters in K-Means clustering analysis [18][19]. This method uses a plot of Inertia (or the sum of squared distances of the samples to their closest cluster center) to the number of clusters, where the optimal number of clusters is identified as the point where changes in Inertia start to slow down [20]. Inertia formula as presented in Equation (5) [21].

$$I = \sum_{i=1}^{k}\sum_{x \in c_i} (x - \mu_i)^2 \tag{5}$$

where:

$k$ : number of clusters

$c_i$ : the set of all data points assigned to cluster i

$\mu_i$ : centroid of cluster i.

b. Silhouette Method

Silhouette method is a method used to evaluate clustering quality by measuring how well each data point is grouped in a particular cluster compared to other clusters [22]. This method calculates the Silhouette score as the ratio between the average distance from each data point to all data points in the same cluster and the average distance from each data point to all data points in the nearest cluster [23]. Mathematically, the Silhouette score can be determined by the formula presented in the following Equation (6) [24].

$$s(x) = \frac{(b(x) - a(x))}{Max(a(x), b(x))} \tag{6}$$

where:

$a(x)$ : the average distance from data point x to all data points in the same cluster

$b(x)$ : the average distance from data point x to all data points in other nearest clusters.

Silhouette scores range between -1 and 1, where a score close to 1 indicates that the data point closely matches the assigned cluster, a score close to 0 indicates that the data point matches both assigned clusters and a score close to -1 indicates that those data points do not match the assigned cluster.

### 2.5 Cluster Analysis with K-Means

Clustering or cluster analysis is a process of organizing a group of unknown data into various groups in such a way that similar objects become one cluster while dissimilar objects become members of another cluster [25]. The method is implemented by determining the desired number of clusters (k), then determining the k centroids of each cluster. Each data point is then assigned to the cluster that has the closest centroid [26]. This process is repeated until there is no change in the data grouping. The steps for analyzing K-Means Clusters are mathematically as follows:

1. Choose the number of k clusters and choose k random points as the centroids of each cluster with the following formula.

$$C = \{c_1, c_2, \dots, c_k\} \tag{7}$$

2. Calculate the distance between each data point and each centroid, using the Euclidean distance formula as follows.

$$d(x,c) = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + \cdots (x_n - c_n)^2} \tag{8}$$

where:

$d(x,c)$ : distance between data and centroid

3. Enter each data point into the cluster that has the closest centroid using the following formula.

$$Cluster(x) = argmin_c\left(d(x,c)\right) \tag{9}$$

4. Calculate the new centroid of each cluster by taking the average of all data points assigned to the cluster with the following formula.

$$c_i = \left(\frac{1}{N_c}\right)\left(\sum_{x \in c} x\right) \tag{10}$$

where:

$x$ : variable

$c$ : centroids

$N_c$ : the amount of data in the cluster c.

5. Repeat step 2 to step 4 such that there are no more changes in the data grouping.

## 3. RESULTS AND DISCUSSION

Writing the results and discussion can be separated into different subs or can also be combined into one sub. The summary of results can be presented in the form of graphs and figures. The results and discussion sections must be free from multiple interpretations. The discussion must answer research problems, support, and defend answers with results, compare with relevant research results, state the study’s limitations, and find novelty.

### 3.1 Data Normalization

To determine the appropriate K-Means method, the initial step of the analysis is to normalize the data. The results of data normalization in this study are an attempt to transform the original variable ($X$) into a standard number ($Z$). Data normalization is used so that the performance of the clustering model can be optimal and each research variable has the same unit.

### 3.2 Assumption Test

Two assumptions must be fulfilled when conducting cluster analysis, namely the sample must be represented using the KMO test (Kaiser Meyer Olkin) and there is no correlation or multicollinearity between variables. On the assumptions of this study, a representative sample uses the KMO value to check the accuracy of a representative sample using the calculation formula in **Equation (2)**. The value obtained is later compared to the value of 50%, where if the KMO value is > 50%, then it is stated that the sample is representative of the population. A summary of the KMO values obtained using the Rstudio Software is shown in **Table 2**.

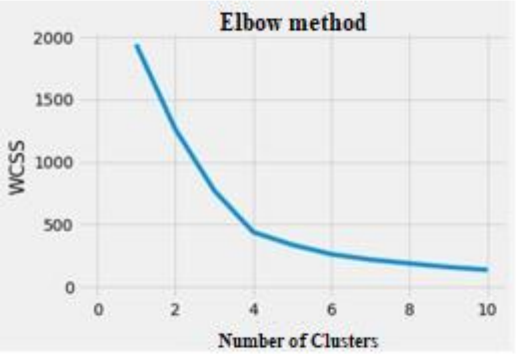
**Table 2. Calculation Results of KMO Values**

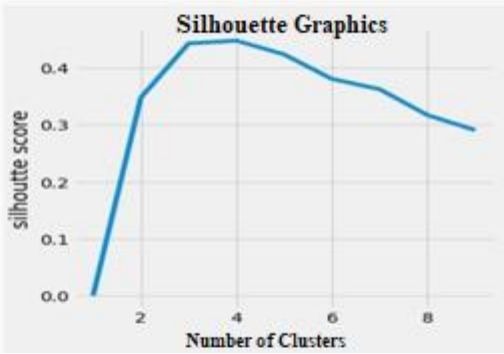
| KMO Value |
|---|
| 0.662833 |

In **Table 2** it can be seen that the KMO value in the data as a whole is 66.2833% and this KMO value is greater than 50% so it is concluded that the research data sample has met the assumptions of a representative sample.

### 3.3 Determine K-Optimal

After the assumptions are met, the next step is to classify the data. The clustering in this study uses the K-Means Algorithm so that the units of observation will be grouped into k groups where k is considered the optimal number of clusters. The limit for determining k is 10 clusters. Then do the calculations based on **Equation (5)** and **Equation (6)**. The optimal number of clusters formed can be determined by the elbow method by looking at the percentage of the results of the comparison between the number of clusters that will form an elbow at a point. The results of this elbow method will be compared with the silhouette method which is used to see the quality and strength of clusters with the average value approach. The higher the average value, the better. Figure 1 shows the optimal cluster with the Elbow method and the Silhouette method.




(a) (b)

**Figure 1. Optimal Cluster Number from Elbow Method (a) and Silhouette Method (b)**

In **Figure 1** it can be seen that the line is broken which forms an elbow when k is 4, while the highest average value on the graph of the Silhouette method also occurs when k = 4. Based on these two methods, the optimal number of clusters is 4 clusters. **Figure 1** presents the results of the cluster plot bar results.

### 3.4 K-Means Analysis

The results of the K-Means analysis divide into 4 clusters, with the cluster plots shown in **Figure 5**.

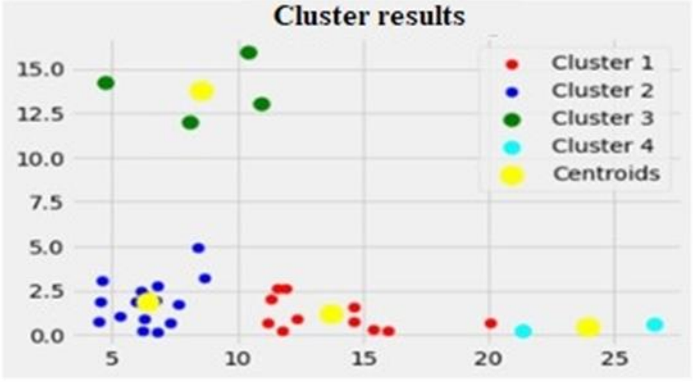



**Figure 2 Plotbar Clustering Results**

When viewed visually, it is clear that the points presented in **Figure 2** have their characteristics where the existing points seem to gather in one place. The K-Means technique in this study uses Euclidean distance where this technique searches for the nearest centroid so that cluster groups are determined from the closest Euclidean distance from the centroid. Visually, it can be seen that there is one object that is included in cluster 1 close to the position in the cluster 4 object, but characteristically the cluster is defined according to the values of the variables as a cluster differentiator, this 1 object is an object that is included in cluster 1. Justification about this object included in cluster 1 has been validated statistically as the values of the variables as cluster differentiators in Table 5, which are based on variables poverty percentage, total transport (vehicles) percentage, transportation use percentage, and territory area percentage.

### 3.5 Cluster Analysis Results Formed

The four clusters are formed based on the condition of the level of subsidy needs based on the characteristics of transportation according to the cluster variables used in this study. Based on the level of need related to transportation subsidies, the four clusters can be interpreted as clusters with low, medium, high, and very high severity. Cluster 1 is a group of provinces with high poverty rates and moderate fuel consumption taking into account the size of the area. Meanwhile, cluster 2 is a province group characterized by a moderate poverty rate and low fuel consumption taking into account the size of the area. Cluster 3 is a province group that has the characteristics of having a low poverty rate and high fuel consumption taking into account the size of the area. Meanwhile, cluster 4 is a province with a high poverty rate and low fuel consumption taking into account its area. **Table 4** presents the members of the four cluster provinces that have been obtained from the analysis.

**Table 4. Results of Province Members for Each Cluster**

| *Clusters* | Provinces | Characteristics |
|---|---|---|
| 1 | Aceh, Sumatra Selatan, Bengkulu, Lampung, DI Yogyakarta, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku | Provinces with high poverty rates and moderate fuel consumption taking into account the size of the area with a total of n=12 provinces |
| 2 | Sumatera Utara, Sumatera Barat, Riau, Jambi, Kep. Bangka Belitung, Kepulauan Riau, Banten, Bali, Kalimantan Barat Kalimantan Tengah, Kalimantan Timur Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, Maluku Utara | Province group characterized by a moderate poverty rate and low fuel consumption taking into account the size of the area with a total of n=16 provinces |
| 3 | Jawa Timur, DKI Jakarta, Jawa Barat, Jawa Tengah | Province group that has the characteristics of having a low poverty rate and high fuel consumption taking into account the size of the area with a total of n=4 provinces |
| 4 | Papua Barat, Papua | Province with a high poverty rate and low fuel consumption taking into account its area |

| *Clusters* | Provinces | Characteristics |
|---|---|---|
| | | with a total of n=2 provinces |

In **Table 4** it can be shown that the members of each cluster are formed, where Cluster 1 has 12 provinces, cluster 2 has 16 provinces, cluster 3 has 4 provinces and cluster 4 has 2 provinces. This shows that different characteristics of each province cluster has been formed based on the 4 variables observed, namely poverty percentage, total transport (vehicles) percentage, transportation use percentage, and territory area percentage. Furthermore, the characteristics of each cluster of the province that was formed were identified. The characteristics of each cluster in the province can be seen through the average of each variable presented in **Table 5**.

**Table 5. Average Calculation Results of Members of Each Cluster in Each Variable**

| *Clusters* | Poverty Percentage | Percentage of Total Transport | Percentage of Transportation Use | Percentage of Territory Area |
|---|---|---|---|---|
| 1 | 13.7033% | 1.1758% | 8.5154% | 1.9550% |
| 2 | 6.3675% | 1.8631% | 9.1916% | 3.0281% |
| 3 | 8.5150% | 13.8075% | 9.4597% | 1.5200% |
| 4 | 23.9450% | 0.4200% | 4.6792% | 11.005% |

The values in **Table 5** show that in terms of poverty, cluster 1 is a province with a high poverty rate. When viewed from the percentage of the number of vehicles and transportation, the areas in Cluster 1 show that the number of vehicles used is relatively moderate in September 2022 when compared to the area, cluster 1 also has a moderate area so Cluster 1 is a province with a high poverty rate and moderate fuel consumption taking into account the size of the area, the density level in cluster 1 is categorized as high, so cluster 1 is highly prioritized for the distribution of fuel subsidies.

In cluster 2, in terms of poverty, it is an area with a low poverty rate with a relatively low percentage of vehicles and transportation in September 2022 when compared to its area which has an area of 3.0281% so cluster 2 is a province with a high poverty rate low and low fuel consumption taking into account the area, then the level of density in cluster 2 is categorized as medium. So that cluster 2 is not too prioritized for the distribution of fuel subsidies.

Cluster 3, in terms of the percentage of the number of vehicles and transportation, is included in the area which is classified as having a very high level of mobility when compared to the area of the area. Meanwhile, from an economic point of view, including provinces that have high poverty rates, it can be stated that the provinces in cluster 3 have high poverty rates and high fuel consumption taking into account their area size, so the density level in cluster 3 is categorized as very high. So cluster 3 is prioritized for the distribution of fuel subsidies. The provincial government of Cluster 3 needs to implement a policy to sort out people who need to get fuel subsidies and those who are not entitled to get fuel subsidies.

In terms of poverty, cluster 4 is an area with a very high poverty rate with a relatively low percentage of vehicles and transportation in September 2022 when compared to its area, so it is assumed that cluster 4 is a province with a high poverty rate and low fuel consumption taking into account the area, the level of density in this group is categorized as low. So cluster 4 is not prioritized for the distribution of fuel subsidies. However, the government can allocate subsidy assistance in the form of food and material needs for Cluster 4.

Based on the cluster results obtained using the K-Means method, it can be projected in the form of a map of Indonesia's territory based on the variables of poverty, percentage of vehicles, transportation, and area. Mapping of provinces in Indonesia as a result of cluster analysis using the K-means method where cluster 1 has 12 provinces, cluster 2 has 16 provinces, cluster 3 has 4 provinces, and cluster 4 has 2 provinces as presented in **Figure 3**.

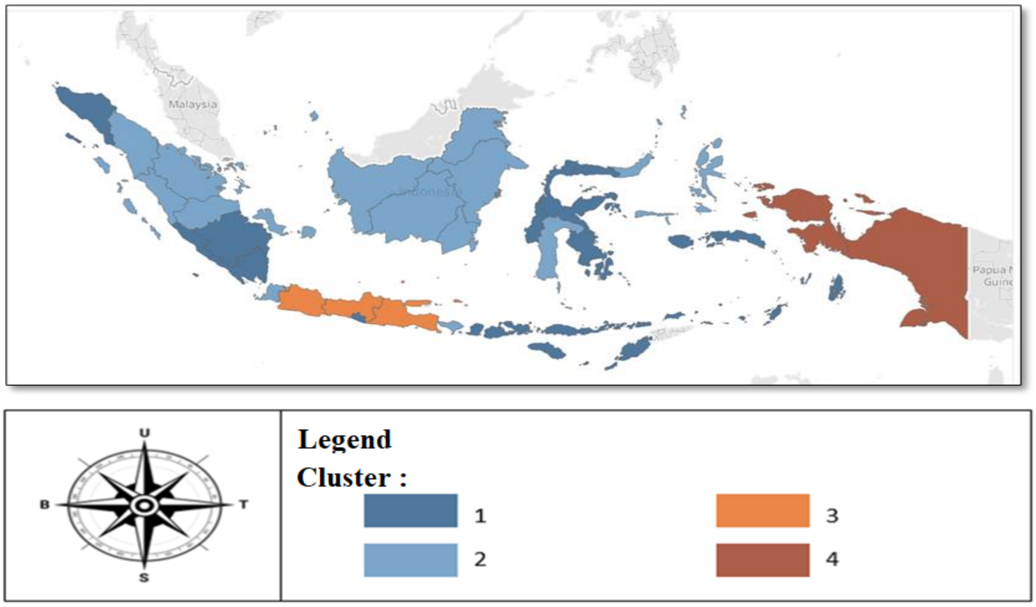



**Figure 3 Map of Indonesia Based on Poverty, Vehicle, and Area Variables with K-Means Method.**

Information:

Cluster 1: Provinces are Highly Prioritized to Receive Fuel Subsidies

Cluster 2: Provinces are not Highly Prioritized for Receiving Fuel Subsidies

Cluster 3: Provinces are Prioritized to Receive Fuel Subsidies

Cluster 4: Provinces are not Prioritized to Receive Fuel Subsidies

## 4. CONCLUSIONS

Based on the results of cluster analysis using K-Means, 4 clusters were formed with each having its characteristics, namely Cluster 1, provinces are highly prioritized to receive fuel subsidies, Cluster 2, provinces are not highly prioritized to receive fuel subsidies, Cluster 3, provinces are prioritized to receive fuel subsidies, and in Cluster 4, provinces are not prioritized to receive fuel subsidies. Indicators of goodness for the formation of clusters based on the Elbow method and the Silhouette method. The results obtained indicate that there is a need for very serious attention regarding areas that have the highest levels of fuel use. This shows that the level of public dependence on the use of fuel is still relatively high. Therefore, the increase in fuel prices makes it difficult for the community because it has an impact on all areas of life which has an impact on rising inflation rates in Indonesia, so the government needs to evaluate this policy to reduce the difficulties felt by the wider community. The results of this study indicate that not all provinces in Indonesia require fuel subsidies, namely the provinces in Cluster 4 and Cluster 3, namely West Papua, Papua, North Sumatra, West Sumatra, Riau, Jambi, and Bangka Belitung Islands, Riau Islands, Banten, Bali, West Kalimantan, Central Kalimantan, East Kalimantan, South Kalimantan, North Kalimantan, North Sulawesi, South Sulawesi, North Maluku. The results of this study provide recommendations to the Government of Indonesia to pay close attention to the provinces in Indonesia that need fuel subsidies in the provincial areas in Cluster 1 and Cluster 3, namely Aceh, South Sumatra, Bengkulu, Lampung, DI Yogyakarta, Nusa Tenggara West, East Nusa Tenggara, Central Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, East Java, DKI Jakarta, West Java, and Central Java.

## ACKNOWLEDGMENT

We would like to thank the data provider institution for this research, all parties who have assisted in the completion of this research, and the anonymous reviewer for this research.

## REFERENCES

[1] E. Wiriani, "Pengaruh Inflasi dan Kurs terhadap Pertumbuhan Ekonomi Indonesia," *J. Samudra Ekon.*, vol. 4, no. 1, pp. 41–

50, 2020.
[2] S. Wulandari and M. D. Habra, “Pengaruh Indeks Harga Konsumen Terhadap Inflasi Di Kota Medan,” in *PROSIDING SEMINAR NASIONAL HASIL PENELITIAN*, 2020, vol. 3, no. 1, pp. 563–568.
[3] BPS, “Dukungan BPS dalam Menekan Kemiskinan Ekstrem,” *Badan Pusat Statistik*, 2022. https://www.bps.go.id/news/2022/01/31/459/dukungan-bps-dalam-menekan-kemiskinan-ekstre (accessed Oct. 20, 2022).
[4] BPS, “Inflasi Maret 2022 Tertinggi Sejak Mei 2019,” *Badan Pusat Statistik*, 2022. https://www.bps.go.id/news/2022/04/01/508/inflasi-maret-2022-tertinggi-sejak-mei-2019.html (accessed Oct. 20, 2022).
[5] W. Harmono, “DAMPAK KEBIJAKAN PENGALIHAN SUBSIDI BBM DI TENGAH KRISIS MULTINASIONAL TERHADAP INFLASI DAN PERTUMBUHAN EKONOMI DI INDONESIA,” *J. Ekon. Kreat. dan Manaj. Bisnis Digit.*, vol. 1, no. 2, pp. 327–333, 2022.
[6] W. Wardani, S. Suriana, S. U. Arfah, Z. Zulaili, and P. S. Lubis, “Dampak kenaikan Bahan Bakar Minyak (BBM) Terhadap Inflasi dan Implikasinya Terhadap Makroekonomi di Indonesia,” *AFoSJ-LAS (All Fields Sci. J. Liaison Acad. Soc.*, vol. 2, no. 3, pp. 63–70, 2022.
[7] R. A. Putri, “DEKONSTRUKSI KEBIJAKAN PUBLIK MASA KINI MELALUI ESKALASI KUALITAS SATU DATA INDONESIA: ANTARA HARAPAN DAN KENYATAAN,” 2022.
[8] C. Chatfield, *Introduction to multivariate analysis*. Routledge, 2018.
[9] S. K. Dini and A. Fauzan, “Clustering provinces in indonesia based on community welfare indicators,” *EKSAKTA J. Sci. Data Anal.*, pp. 56–63, 2020.
[10] M. Nusrang, M. K. Aidid, and Z. Rais, “K-Means Cluster Analysis for Grouping Districts in South Sulawesi Province Based on Village Potential,” *ARRUS J. Math. Appl. Sci.*, vol. 2, no. 2, pp. 73–82, 2022.
[11] A. Supriyadi, A. Triayudi, and I. D. Sholihati, “Perbandingan algoritma k-means dengan k-medoids pada pengelompokan armada kendaraan truk berdasarkan produktivitas,” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 6, no. 2, pp. 229–240, 2021.
[12] C. Beckett, L. Eriksson, E. Johansson, and C. Wikström, “Multivariate data analysis (MVDA),” *Pharm. Qual. by Des. A Pract. approach*, pp. 201–225, 2018.
[13] U. Braga-Neto, *Fundamentals of pattern recognition and machine learning*. Springer, 2020.
[14] T. Niu *et al.*, “Study of degradation of fuel cell stack based on the collected high-dimensional data and clustering algorithms calculations,” *Energy AI*, vol. 10, p. 100184, 2022.
[15] X. Pan, M. L. Song, J. Zhang, and G. Zhou, “Innovation network, technological learning and innovation performance of high-tech cluster enterprises,” *J. Knowl. Manag.*, vol. 23, no. 9, pp. 1729–1746, 2019.
[16] A. Afifi, S. May, R. Donatello, and V. A. Clark, *Practical multivariate analysis*. CRC Press, 2019.
[17] J. Costales, J. J. J. Catulay, J. Costales, and N. Bermudez, “Kaiser-Meyer-Olkin Factor Analysis: A Quantitative Approach on Mobile Gaming Addiction using Random Forest Classifier,” in *Proceedings of the 6th International Conference on Information System and Data Mining*, 2022, pp. 18–24.
[18] P. Fränti and S. Sieranoja, “K-means properties on six clustering benchmark datasets,” *Appl. Intell.*, vol. 48, pp. 4743–4759, 2018.
[19] S. Bera, D. Chakrabarty, N. Flores, and M. Negahbani, “Fair algorithms for clustering,” *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.
[20] A. Lengyel and Z. Botta-Dukát, “Silhouette width using generalized mean—A flexible method for assessing clustering efficiency,” *Ecol. Evol.*, vol. 9, no. 23, pp. 13231–13243, 2019.
[21] Y. P. R. Álvarez, S. D. L. O. Guerra, and J. J. F. Díaz, “CLUSTER ANALYSIS FROM A RESEARCH STUDY ON DIGITAL COMPETENCES IN UNIVERSITY PROFESSORS,” *PalArch’s J. Archaeol. Egypt/Egyptology*, vol. 18, no. 3, pp. 4888–4911, 2021.
[22] C. Maione, D. R. Nelson, and R. M. Barbosa, “Research on social data by means of cluster analysis,” *Appl. Comput. Informatics*, vol. 15, no. 2, pp. 153–162, 2019.
[23] P. Govender and V. Sivakumar, “Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019),” *Atmos. Pollut. Res.*, vol. 11, no. 1, pp. 40–56, 2020.
[24] D. M. SAPUTRA, D. SAPUTRA, and L. D. OSWARI, “Effect of distance metrics in determining k-value in k-means clustering using elbow and silhouette method,” in *Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*, 2020, pp. 341–346.
[25] M. R. Yudhanegara, S. W. Indratno, and R. R. K. N. Sari, “Clustering for Item Delivery Using Rule-K-Means,” *J. Indones. Math. Soc.*, vol. 26, no. 2, pp. 185–191, 2020.
[26] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, “Integration k-means clustering method and elbow method for identification of the best customer profile cluster,” in *IOP conference series: materials science and engineering*, 2018, vol. 336, p. 12017.