UAD Universitas Ahmad Dahlan

Indonesian provincial clustering using Elbow method for the national food security during pandemic

Rido Trimanto[1], Erykka Yustari[1], Zulfatin Nafizah[1], Nona Carolina[1], Nursyiva Irsalinda [1*], Arifah Indah Setyorini[2]

[1]Universitas Ahmad Dahlan, Jl. Jend. Ahmad Yani, Tamanan, Banguntapan, Bantul, DIY 55191, Indonesia
[2]Universitas Negeri Yogyakarta, Jl. Kolombo No. 1, Karangmalang, Sleman, DIY 55281, Indonesia

*Corresponding e-mail: nursyiva.irsalinda@math.uad.ac.id

ARTICLE INFO

ABSTRACT

The Covid-19 pandemic had an impact on the joints of socio-economic life, especially in fulfilling the basic needs. It also caused the declining of global food security, especially in Indonesia. This study aims to develop regional mapping to determine food security priorities and to achieve equal distribution of food security throughout Indonesia. The research method used in this research is quantitative research with the Elbow method. The Elbow method is used to find the optimal cluster size. The data used are from the Food Security Agency of the Indonesian Ministry of Agriculture and Central Statistics Agency in a range of 2020 to 2021. In the process to identify priority areas in Indonesia, it is necessary to have provincial clustering. It is also necessary to minimize food budget allocations that are not well-targeted, causing losses, and not achieving an equal distribution of food security programs. Looking from a more visionary perspective, the success of clustering provides an opportunity for the government to focus more on allocating budget, resources, and time according to the results of the clustering. Based on the results of the provincial clustering, two clusters were obtained, namely provinces with high food security (Cluster 1) and low food security (Cluster 2). Cluster 1 has lower constituent components than Cluster 2.

Introduction

Coronavirus disease 2019 (Covid -19) is a positive single strain RNA virus, encapsulated and not segmented (Yuliana, 2020). Covid -19 has been declared by the WHO as a pandemic and the Government of Indonesia based on Presidential Decree No. 11 of 2020 concerning the Establishment of a Covid -19 Public Health Emergency that must be taken care of (Syauqi, 2020). The Covid -19 pandemic that occurred in Indonesia had an impact on the joints of socio-economic life, especially in meeting basic needs (Masniadi et al., 2020). The government's policy towards

people to work from home and maintain physical distance as well as several government regions that have implemented policies to partially quarantine the area have changed the situation, including changes to new food supply chain patterns in almost all aspects of life. The work pattern or system in the food sector does seem to have changed very significantly starting from the production process to consumption from upstream to downstream during the Covid-19 pandemic (Hirawan & Verselita, 2020).

In 2019 around 8.9% of the total world population is currently experiencing hunger (Ayinde, et al., 2020). As a result of the Covid -19 pandemic, various countries experienced a recession which increased the number of acute hunger. With this increase, it is estimated that 83-132 million people will experience hunger (Klassen & Murphy, 2020). Covid -19 also causes employment in agriculture to contract by 4.87 percent, while domestic agricultural production will shrink by 6.2 percent. Imports will decrease by 17.11 percent and prices are expected to increase by 1.20 percent in the short term and by 2.42 percent in 2022. With reduced domestic supply and imports, food shortages and food price inflation have the potential to occur (Asmanto et al., 2020). Not only that, but Covid -19 has also caused the wheels of the economy to stop due to regulations for the implementation of community activity restrictions. Indirectly, this also causes global food security to decline drastically. In addition, the restriction is also an obstacle in the transportation distribution of food stocks in various regions. Another severe impact during the Covid -19 pandemic is the unsold or non-delivery of crops to consumers. Based on the description of the problems above, it can be concluded that food availability is an aspect of national food security that needs to be considered by the government. This is because food is the most basic need besides housing and clothing.

The United Nations World Food Security Committee defines food security as a situation in which all people and under all circumstances have physical, social, and economic access to food, which must not only be sufficient but also safe and nutritious (Wulandari & Anggraini, 2020). Based on Government Regulation of the Republic of Indonesia Number 17 of 2015 concerning Nutrition and Food Security, food security is a condition of fulfilling food for the state to individuals, which is reflected in the availability of sufficient food, then in terms of quality and quantity, nutritious, safe, equitable, diverse, and affordable. In addition, the food must not conflict with the religion and culture of the community. Food availability is an important prerequisite for sustainable consumption but is considered inadequate in the context of food security because many factors affect the achievement of food security at both the household and regional levels. Observing current conditions, especially the pandemic situation due to Covid-19 which affects important aspects of people's lives, especially in the economic field, namely food.

Determining the region as a priority for food security is one of the government's efforts in distributing food security in Indonesia during the Covid-19 pandemic. Therefore, smart and preventive steps are needed to carry out provincial clusters to realize equitable distribution of food security and identify areas that are priority areas in increasing food security in efforts to reduce hunger in Indonesia. This is necessary to minimize food budget allocations that are not well-targeted, causing losses and not achieving an equitable distribution of food security programs. From a more visionary perspective, the success of clustering provides an opportunity for the government to focus more on allocating budget, resources, and time according to the results of the clustering. So the purpose of writing this article is to present data on the clustering of provinces in Indonesia using the Elbow method as a solution to national food security during a pandemic.

Method

The research method used in this study is the Elbow method. The Elbow method is a method for generating information by looking at the comparison of the results between the number of clusters that will form a final point, then the value cluster will be used as a data model for the best cluster. In addition, the calculation results will be used to compare the number of clusters. The type of

research conducted in this research is quantitative research, which analyzes numerical data (numbers). In the research methodology, there are steps of research procedures, starting, literature study, data collection, implementation of the K-Means algorithm and Hierarchical clustering algorithm, as well as testing the Elbow method.

First, the author conducted a literature study through several reading sources. Second, the authors collected secondary data released by the Food Security Agency of the Ministry of Agriculture. The data represents the variables of food affordability, food availability, and food utilization. Then for the data on the variable area of food land, the author takes from the Central Statistics Agency. Third, before the data is divided into several groups, the author uses the Elbow method to test the number of the best clusters to be used.

In the last step, the author performs the first data processing using the K-Means method which aims to cluster provinces that have the lowest to the highest levels of food security. The K-Means method is a clustering method that divides the data into k clusters with the closest average. The second data processing using Hierarchical clustering has the same purpose as the K-Means method with the concept of data grouping methods starting with grouping more than two objects that have the closest resemblance. The results of the two methods are then compared to get the best model.

Results and Discussion

Clustering is a grouping technique based on the similarity of data characteristics so that it can facilitate decision-making in a policy. The same thing is also explained by Metisen and Sari (2015), clustering is a method used to divide a data series into several groups based on predetermined similarities so that objects in one cluster will have a high similarity between each other (Aulia, 2020). Based on the explanation that has been described, the research carried out is to use the clustering technique to cluster provinces in all regions in Indonesia to consider priority areas so that food security is evenly distributed.

There are three basics in food security, namely availability, affordability, and utilization of food. The three basics have different roles in determining food security clusters, such as food readiness which describes the existence of food production, transactions, and food reserves, food affordability which describes the ability to obtain sufficient and nutritious food for the benefit of households, and aspects of food utilization (Pinstrup-Andersen, 2009). In addition, the existence of agricultural land can have great benefits for the social, economic, and environmental fields. The decline in agricultural land has resulted in a decrease in food products so the land area is quite influential on food security (Destianto & Pigawati, 2014). An area is declared successful in developing food security if there is an increase in food production, smooth food distribution, and safe and nutritious food consumption for the entire community. The basic parameters of food security can be seen in Table 1.

Based on Table 1, an area is categorized as having good food security if the indicators of the three pillars are met. To achieve the goal of equitable distribution of food security, appropriate techniques or planning are needed.

Table 1. Basic parameters of food security

Base	Parameter
Food Availability	- Relationship of per capita consumption with food production.
Food Affordability	- Percentage of population below the poverty line.
	- The percentage of households is more than 65% with the proportion of food expenditure to total expenditure.
	- Percentage of households without electricity.
Food Utilization	- The average school year for girls is more than 15 years.
	- Proportion of households without access to clean water.
	- Percentage of population per health worker to density level.
	- Prevalence of stunting under five.
	- Life expectancy at birth.

The initial strategy that must be carried out is mapping the problems and solutions of national food security in Indonesia related to food security problems. Various aspects that affect food security are divided into 4 variables, including food affordability, food availability, food utilization, and food area. These variables become the potential differentiator for food security in each province in Indonesia so that it can be known which provinces should be prioritized inequitable distribution of food security. These variables are then clustered to obtain priority provincial groups. The results of the clustering obtained can be used as a guide for the government in the policy-making process related to food security issues. Then the mapping of food security problems and solutions that have been described can be seen in Figure 1.

Figure 1. Mapping of national food security problems and solutions in Indonesia

In this study, the authors emphasize three fundamental values that must be considered in improving food security in Indonesia. The first fundamental value is that food security must benefit the community, the second prioritizes the provinces that have the lowest cluster level in terms of food security, and the third is to provide a budget that is in accordance with the cluster level of a region. Regions in the lowest cluster are given a larger budget than the other clusters. Factors considered in determining the priority of this model are the variables of food affordability, food availability, food utilization, and land area.

This study uses secondary data released by the Food Security Agency of the Ministry of Agriculture. The data represents the variables of food affordability, food availability, and food utilization. Meanwhile, the variable regarding the area of food land is based on data released by the Central Statistics Agency. These variables become the differentiating provinces in Indonesia so that it can be obtained which provinces are the priority in increasing food security. These variables also become features in determining provincial clusters where each cluster has a different proportion of handling. In this study, the clustering method was used to group 34 provinces in Indonesia based on the level of food security. Comprehensive, accurate, and periodic data on resilience is needed to improve food security and face food insecurity in Indonesia.

The analysis in this research data uses data mining by implementing certain processes. Data mining is a collection of large amounts of data that is processed to obtain information, interesting data patterns, and some knowledge (Sari et al., 2022). In data mining, several techniques can be used, one of which is clustering, which has been mentioned in the previous description. According to Sucipto (2019), the notion of clustering in data mining is a technique for grouping data into a certain cluster that allows the data in that cluster to have similarities and have clear differences with data in other clusters.

In this study, the authors compared two clustering techniques, namely K-Means clustering and Hierarchical clustering. Hierarchical clustering is a data grouping method that begins with grouping more than two objects that have the closest resemblance. The next process is by

grouping the proximity of the two other objects that have the closest distance. This process continues until the cluster will form a tree in which there is a clear hierarchy (level) between objects from those that have similarities to those that have the least resemblance (Khormarudin, 2016). Meanwhile, K-Means clustering is one of the clustering methods (unsupervised), where the data is divided into one or more clusters. In this case, K-Means clustering models the dataset into groups so that data items that experience clusters have different characteristics from the group and have other similar characteristics (Wakhidah, 2014). The purpose of the grouping is to minimize the objective function in the grouping process to maximize variation between groups and minimize variations within a group. The results of the clustering are provincial clusters with the highest to the lowest priority based on the four variables involved.

Then the authors do clustering based on the variables of food availability, food affordability, food utilization, and land area. Before performing clustering, it is necessary to determine the best number of clusters, the author uses the Elbow method to determine the number of clusters. Based on the results of cluster determination using the Elbow method, the results obtained are presented in Figure 2.
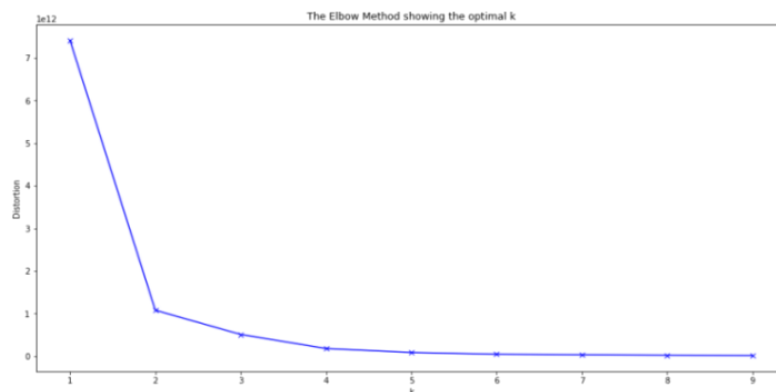

Figure 2. Elbow chart

From Figure 2, it can be seen that the graph fracture is very significant in the cluster with several 2, so 2 is the number of the best clusters. Then after determining the number of the best clusters, provincial clustering is carried out. The first clustering uses Hierarchical clustering. The first stage is data normalization. It aims to make all variables in the same range. Then the similarity distance is calculated using the agglomerative method to obtain the dendrogram shown in Figure 3.
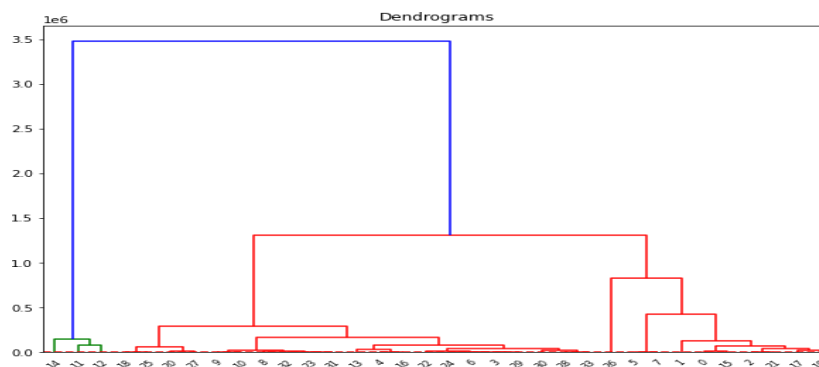

Figure 3. Dendrogram of hierarchical clustering

From Figure 3, it can be seen that the members of cluster 1 are West Sumatra, Riau, Jambi, Aceh, North Sumatra, Bengkulu, Lampung, Bangka Belitung Islands, Riau Islands, DKI Jakarta, West Java, DIY, Central Java, East Java, Banten, Bali, NTB, West Sulawesi, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Central

Sulawesi, Southeast Sulawesi, South Sulawesi, Gorontalo, South Sumatra and North Maluku. Cluster 2 members are Maluku, West Papua, East Nusa Tenggara, and Papua.

From the results of clustering, it can be concluded that cluster 1 is a province with higher food security than cluster 2 which has lower food security. The second technique is using K-Means. The first stage is carried out by normalizing the data on each variable, the next stage is determining the number of clusters from 34 provinces in Indonesia from the Elbow method, the best number of clusters used is 2. The third stage is determining the centroid value in each cluster. The next process calculates the distance between the centroid point and the fixed points. Finally, each object is allocated to each centroid so that the grouping results are obtained which are shown through the scatter plot in Figure 4.
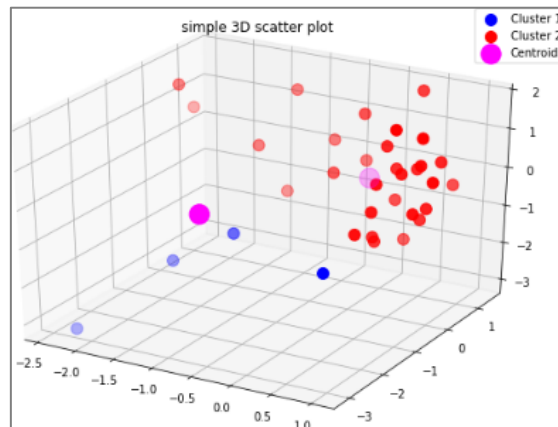

Figure 4. 3D scatter plot K-Means

From Figure 4, it can be seen that the division of data points into clusters is appropriate, where it can be seen that all data points are included in their respective clusters. In addition, the division of clusters looks good, this can be seen from the distance between the members of the cluster who are not close to each other. Based on the results of clustering using the two methods above, the accuracy of each method is shown in Table 2.

Table 2. Clustering Method Comparison

	Clustering Method	Silhouette Score	DBI
0	K-Means	0.401874	0.876456
1	Hierarchical	0.401874	0.876456

In Table 2, there is a silhouette score for the Hierarchical clustering method and the K-Means clustering method has the same value, namely 0.401874. Based on (Ogbuabor & F. N, 2018) if the silhouette score is closer to 1 then the object grouping is getting better, while if the silhouette score is close to -1 then the objects are not grouping properly. From the provisions of the silhouette score, it can be concluded that the value of the silhouette score using the Hierarchical clustering and K-Means clustering methods shows that the objects are grouped correctly. Then the Davies-Bouldin Index (DBI) value is one of the internal evaluation methods to measure cluster evaluation in the grouping method based on the values of cohesion and separation. If the DBI value is small or obtained (non-negative >= 0), then the cluster results obtained are getting better. Based on the results of Table 2, there is a Davies-Bouldin Index (DBI) of the two methods above of 0.876456. Since both methods have the same silhouette score and Davies-Bouldin Index, one of them can be used. The results of grouping the selected methods are depicted in Figure 5.

Based on the results of the provincial food security cluster using the K-Means method, two clusters were obtained, namely high and low. For provinces that are included in the high cluster, there are 29 provinces, namely West Sumatra, Riau, Jambi, Aceh, North Sumatra, Bengkulu,

Lampung, Bangka Belitung Islands, Riau Islands, DKI Jakarta, West Java, DIY, Central Java, East Java, Banten, Bali, NTB, West Sulawesi, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Central Sulawesi, Southeast Sulawesi, South Sulawesi, Gorontalo, South Sumatra and North Maluku.
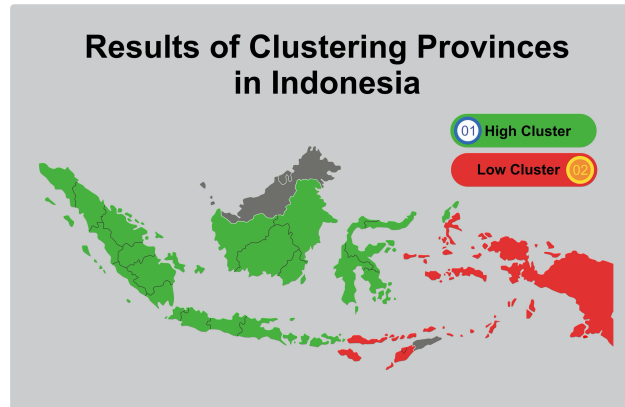

Figure 5. Results of clustering provinces in Indonesia

Meanwhile, those in the lower cluster are Maluku, West Papua, East Nusa Tenggara, and Papua. Because there are different levels of food security in each province, it is necessary to have different policies related to the food system. Provinces with low levels of resilience are prioritized in handling. Although all provinces experience food insecurity due to the impact of Covid-19, provinces with low levels of food security will have the most massive impact. The existence of different conditions in each province requires a differentiation strategy. The strategies are grouped based on aspects of food security, including food availability, food access, and food utilization. The national food security strategy in dealing with Covid-19 is seen in Table 3.

Table 3. Planning in National Food Security

Food Resistance Level	Food Availability	Food Affordability	Food Utilization
Low	- Increase the number of food barns as food reserves. - Improving post-harvest technology to improve efficiency in product processing, packaging, storage, and transportation.	- Improving aspects of distribution and transportation to provide fair and equitable food. - Maintaining food price stability.	- Improving technology to increase the level of food productivity and diversification in food quality and types of food. - Generate economic incentives to support the power of agribusiness. - Improving the quality of food circulating in the community.
High	- Guarantee the amount of food, the goal is that there is no scarcity in society.	- Maintain food price stability.	- Ensuring the quality of food circulating in the community.

Conclusion

Overall, it can be concluded that preventive efforts in overcoming food security are to prioritize provincial areas with lower food security so that in the future it is hoped that there will be equitable distribution of food security in all regions in Indonesia. With the results of the clustering that has been carried out, the government is expected to be able to allocate a budget that is by the results of the clustering of food security. The application of statistics and data science can offer solutions by creating priority clusters based on variables that are considered to have urgency.

These variables include food availability, food affordability, food utilization, and land area. Although there are limitations in the clustering of priority provinces, it is hoped that this method can be adopted on a smaller scale such as in districts or cities.

Acknowledgement

References

Asmanto, P., Adji, A., & Sutikno. (2020). *Menjaga Ketahanan Pangan di Tengah Pandemi Covid-19*. Sekretariat Tim Nasional Percepatan Penanggulangan Kemiskinan. http://tnp2k.go.id/download/92111PB8%20Ketahanan%20PanganFA-Jul2020.pdf.

Aulia, S. (2020). Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja). *Djtechno: Jurnal Teknologi Informasi, 1*(1), 1-5.

Ayinde, I. A., Otekunrin, O. A., Akinbode, S. O., & Otekunrin, O. A. (2020). Food security in Nigeria: impetus for growth and development. *Journal of Agricultural Economics and Rural Development*, *6*(2), 808-820.

Destianto, R., & Pigawati, B. (2014). Analisis keterkaitan perubahan lahan pertanian terhadap ketahanan pangan Kabupaten Magelang berbasis model Spatio Temporal Sig. *Geoplanning: Journal of Geomatics and Planning, 1*(1), 21–32.

Hirawan, F. B., & Verselita, A. A. (2020). *Kebijakan pangan di masa pandemi Covid-19* (pp. 1-7). CSIS Indonesia.

Khormarudin, A. N. (2016). Teknik data mining: Algoritma K-Means clustering. *Jurnal Ilmu Komputer*, 1–12.

Klassen, S., & Murphy, S. (2020). Equity as both a means and an end: Lessons for resilient food systems from COVID-19. *World Development, 136*, 105104.

Masniadi, R., Angkasa, M. A. Z., Karmeli, E., & Esabella, S. (2020). Telaah kritis ketahanan pangan Kabupaten Sumbawa dalam menghadapi pandemi Covid-19. *Indonesian Journal of Social Sciences and Humanities, 1*(2), 109-120.

Metisen, B. M., & Sari, H. L. (2015). Analisis clustering menggunakan metode K-Means dalam pengelompokkan penjualan produk pada Swalayan Fadhila. *Jurnal Media Infotama, 11*(2).

Ogbuabor, G., & Ugwoke, F. N. (2018). Clustering algorithm for a healthcare dataset using silhouette score value. *International Journal of Computer Science & Information Technology, 102*(2), 27-37.

Pinstrup-Andersen, P. (2009). Food security: definition and measurement. *Food Security, 1*(1), 5–7.

Sari, R. Y., Oktavianto, H., & Sulistyo, H. W. (2022). Algoritma K-Means Dengan Metode Elbow Untuk Mengelompokkan Kabupaten/Kota Di Jawa Tengah Berdasarkan Komponen Pembentuk Indeks Pembangunan Manusia. *Jurnal Smart Teknologi, 3*(2), 104-108.

Sucipto, A. (2019). Klasterisasi Calon Mahasiswa Baru Menggunakan Algoritma K-Means. *Jurnal Science Tech, 5*(2), 50-56.

Syauqi, A. (2020). Jalan Panjang COVID-19 (Sebuah refleksi dikala wabah merajalela berdampak pada perokonomian). *JKUBS, 1*(1), 1-19.

Wakhidah, N. (2010). Clustering menggunakan k-means algorithm. *Jurnal Transformatika, 8*(1), 33-39.

Wulandani, B. R. D., & Anggraini, W. (2020). Food estate sebagai ketahanan pangan di tengah pandemi covid-19 di Desa Wanasaba. *SELAPARANG Jurnal Pengabdian Masyarakat Berkemajuan, 4*(1), 386-390.

Yuliana (2020). Corona virus diseases (Covid-19): Sebuah tinjauan literatur. *Wellness and Healthy Magazine, 2*(1), 187-192.