ENTHUSIASTIC
INTERNATIONAL JOURNAL OF APPLIED STATISTICS AND DATA SCIENCE
Volume 4, Issue 1, April 2024, pp. 1-12

K-Means Clustering Application of Open Unemployment in 2020 Caused by COVID-19 in West Java Province

M. Ficky Haris Ardiansyah [a,1,*], Nurfatimah Amany [b,2], Cahya Ireno Anugrah [a,3], Utami Dyah Syafitri [a,4]
[a] Department of Statistics, IPB University, Kampus IPB, Jl. Raya Dramaga, Babakan, Kec. Dramaga, Kabupaten Bogor, 16680, Indonesia
[b] Department of Mathematics, IPB University, Kampus IPB, Jl. Raya Dramaga, Babakan, Kec. Dramaga, Kabupaten Bogor, 16680, Indonesia
[1] fickyardiansyah@apps.ipb.ac.id *; [2] amanynurfatimah@apps.ipb.ac.id; [3] cahyairenocahya@apps.ipb.ac.id; [4] utamids@apps.ipb.ac.id
* Corresponding author

ARTICLE INFO

Keywords
k-means clustering
open unemployment
West Java Province

ABSTRACT

West Java was the province with the highest unemployed rate during the COVID-19 pandemic. Significant increase of open unemployment rate in West Java negatively impacts the national income. This study aims to apply the clustering method using the k-means algorithm to determine priority clusters in West Java Province by looking at the number of clusters in West Java's city and the main characteristic of each cluster. The clustering was conducted utilizing a k-means clustering algorithm which is grouping data based on similar characteristics. The clustering results were evaluated using silhouette method. The results indicated that two clusters were optimal. The clustering process using the k-means method showed that there were three clusters distinguishing the open unemployment rate during the pandemic in West Java Province in 2020. Cluster 1 had a fairly low open unemployment rate due to the stalled service sector and low minimum city wage. Cluster 2 had a high open unemployment rate due to the service sector and high minimum city wage. Cluster 3 had medium open unemployment rate due to the service sector and also medium minimum city wage. It suggests that cluster 2 is a priority cluster in dealing with the open unemployment rate.

1. Introduction

West Java was the province with the highest unemployment rate during the COVID-19 pandemic era [1]. According to the Ministry of Manpower of the Republic of Indonesia (Kementerian Ketenagakerjaan/Kemenaker) of the Republic of Indonesia, in West Java Province, 2.53 million people were unemployed, which was about 25.93% of total unemployment in Indonesia [2]. The COVID-19 pandemic has contributed to the increase in unemployment due to company closure [3]. At least one household member lost their job in more than one in every two households during the pandemic era [4].

Significant increase of open unemployment rate in West Java negatively impacts the national income and public welfare. From 2016 until 2019, West Java was always the third

p-ISSN 2798-253X
e-ISSN 2798-3153

most contributing province in Gross Domestic Product (GDP) Indonesia, after DKI Jakarta and East Java [5]. When the most contributing provinces collapse, the national economy will also collapse. On the other hand, increasing unemployment rate drives poverty and criminality rate besides blocking long term economic growth [6]. This condition has worsened the COVID-19 pandemic's impacts.

The government must conduct some interventions must overcome the unemployment rate in West Java. However, there are many cities which have their own characteristics. To make the intervention efficient, some cities must be designed as priority cities. In determining these priority cities, the k-means clustering was utilized. K-means clustering algorithm is grouping data based on similar characteristics. The cities in a cluster as results of this study are expected to be the priority area for addressing unemployment problems.

2. Method

2.1 Literature Review

Clustering technique is one of many techniques in data mining. Clustering is defined as grouping a set of data or objects into a cluster (group), so each cluster has a similar characteristic but is different from other clusters [7]. The objects are grouped based on their characteristics which are similar to each other. To see how close each object to each other, Euclidean distance was used. Euclidean distance is given by (1).

$$d_{ij} = \left[(x_i - x_j)'(x_i - x_j)\right]^{\frac{1}{2}} \quad (1)$$

where d_{ij} is the distance between the *i*th object and *j*th object, x_i is a variable in vector form for *i*th object, and x_j is a variable in vector form for *j*th object [8].

There are two conventional methods to determine clusters: hierarchical and nonhierarchical. Hierarchical method is used to group objects in a structured way based on their similar characteristics where the number of the clusters is unknown [9]. There are two procedures in the hierarchical method, namely the agglomerative procedure and the divisive procedure [8]. Measuring the similarity between clusters can use the single link, complete link, centroid link, mean link, median link, and ward link [10].

The next method is nonhierarchical. The nonhierarchical method is used for grouping objects where the number of clusters is previously determined as part of the clustering procedure [8]. This method also can be applied to larger data than the hierarchical method.

One example of nonhierarchical method is the k-means algorithm. K-means is a clustering algorithm that divides each data item into one cluster. The following are the steps in the k-means algorithm [11].

a. The number of clusters (k) are determined in the dataset.
b. Centroids for each cluster are determined. First, the centroid points are randomly determined before the iteration to evaluate is done. The formula in iteration procedure is:

$$V_k = \frac{\sum_{i=1}^{N_i} X_i}{N_k} \quad (2)$$

where V_k is the centroid in *k*th cluster, X_i is the *i*th data, and N_k is the number of object/data in cluster k.
c. The closest distance between centroid and each object for each iteration must be calculated using the Euclidean distance formula.

p-ISSN 2798-253X
e-ISSN 2798-3153

2.2 Related Works

This research used the k-means clustering method to group cities based on open unemployment caused by the COVID-19 pandemic in West Java because. K-means clustering method is selected since it is nonhierarchical method and can be used in larger data [10]. It is considered more suitable for this research than other nonhierarchical clustering, such as k-medoids.

There has been numerous research that use the k-means method. This previous research are the base of this research using k-means clustering. A study conducted by Fitriadi and Kurniawati concluded that the k-means clustering method had better performance when compared to the k-medoids algorithm [12]. The study clustered the employee performance data at a national housing company. K-means clustering method is better than k-medoid because it has a higher level of accuracy and returns when compared to the k-medoids algorithm [12]. A study conducted by Sibuea and Sapta mapped outstanding students using the k-means method [13]. It was concluded that the accuracy of the data was quite good for problems related to student achievement with the results of 70%, allowing for recognizing the data used as samples [13]. Nhita analyzed the time performance needed of k-means and k-medoids using High-Performance Computing (HPC) [14]. Parallelized algorithms k-means and k-medoids used the Message Passing Interface (MPI) library. Results research showed that the k-means algorithm yielded smaller Sum of Squares the Error (SSE) than that of k-medoids. The parallel algorithm using MPI provides faster computing [14]. Ramadhani and Januarita compared evaluation results on k-means and k-medoids using a small dataset namely the Iris and Wine datasets [15]. The Davies Bouldin Index was used in this study to measure the similarity of the cluster size based on the spread of data on the cluster and cluster size dissimilarity. Based on the experimental results, k-means showed better evaluation results compared to k-medoids in handling small datasets. It is indicated by the results of the evaluation on the Iris dataset using k-means, which was equal to 0.662, while the Wine dataset showed the results evaluation of 0.534 [15].

2.3 Data

The data for this research were a secondary data obtained from Statistics Indonesia of West Java through National Labor Force Survey (Survei Angkatan Kerja Nasional/SAKERNAS) in August 2020. This dataset contained 59,680 entries. Additive variable for this research was obtained from the decree of West Java Governor on city minimum wage in West Java Province.

2.4 Research Variable

The research variables are presented in Table 1.

Table 1. Units of Research Variables

Variable	Description
Percentage of last education high school or higher	For each city in West Java, the percentage of respondents who have senior high school, diploma, bachelor, master, or doctor from all the respondents.
Percentage of temporary layoff because of COVID-19	For each city in West Java, the percentage of respondents who have a temporary layoff because of COVID-19 from the respondents who have a temporary layoff.
Percentage of permanent layoff because of fired	For each city in West Java, the percentage of respondents have permanent layoff because of firing from the respondents who are fired.
Percentage of permanent layoff because of COVID-19	For each city in West Java, the percentage of respondents who get permanent layoff because of COVID-19 from the respondents who had permanent layoff.

p-ISSN 2798-253X
e-ISSN 2798-3153

Variable	Description
Minimum city wage	Each city's minimum wage in West Java in 2020.

2.5 Research Stage

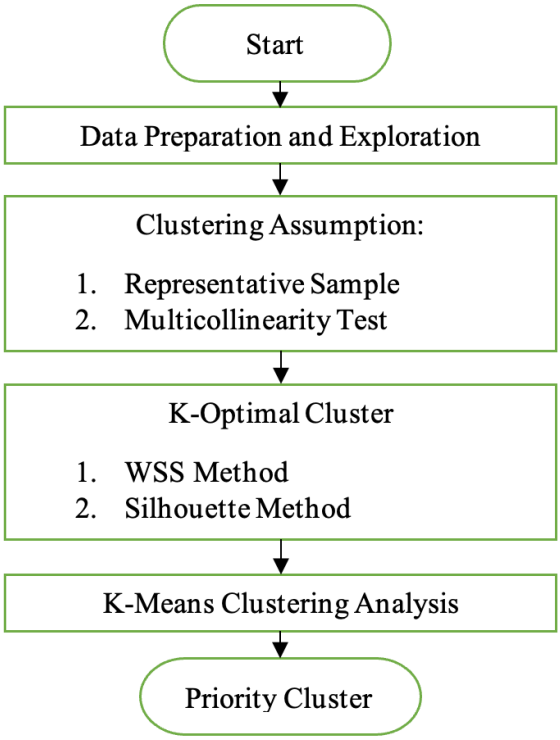

Fig. 1 Research stage.

This research used k-means algorithm with the help of R-studio. This research aims to group regencies in West Java by factors determining open unemployment cases in 2020. Fig. 1 presents the step in this research.

2.6 Cluster Analysis Assumption

2.6.1 Kaiser-Mayer Olkin Test

The Kaiser-Mayer Olkin (KMO) test is used to see whether the sample is sufficiently representative of the existing population, so that the clustering can be processed correctly. This KMO test measures the adequacy of the sample for each indicator. The KMO test has a value of 0 to 1. If the KMO value ranges from 0.5 to 1, the sample can represent the population or a representative sample. KMO is a comparison index of correlation coefficient value to partial correlation [16]. The KMO formula is shown in (3).

$$KMO = \frac{\Sigma_{i=1}^{n}\Sigma_{j=1}^{n} r_{ij}^2}{\Sigma_{i=1}^{n}\Sigma_{j=1}^{n} r_{ij}^2 + \Sigma_{i=1}^{n}\Sigma_{j=1}^{n} a_{ij}^2} \quad (3)$$

2.6.2 No Multicollinearity

Multicollinearity is a perfect or definite linear relationship between several variables or more. It refers to the presence of more than one definite linear relationship. One way to determine the existence of multicollinearity is by looking at the magnitude of the correlation between the independent variables [17]. Multicollinearity can be seen from the value of Variance Inflation Factor (VIF).

$$VIF_i = \frac{1}{1-R_i^2} \quad (4)$$

2.7 K-Means Clustering

K-means is an algorithm for object clusters by attribute to k partition, where k < n. In general, k-means clustering is one of the clustering in non-hierarchical methods. It works by grouping data in the form of one or more clusters or groups. This method partitions data into cluster so that data that have the same characteristics are grouped into the same cluster and every data in a cluster have different characteristics into another cluster.

K-means clustering:

a. Determine the number of clusters using the elbow method with the silhouette method.
b. Determine the centroid value. In determining the centroid value for start of iteration, initial value of centroid is done randomly. Whereas if specifying the value of the centroid which is the stage of the iteration, then the following formula is used:
   1) Calculating the distance between points centroid with point of each object.
   2) Object grouping for determining the cluster member is considering the distance minimum object.
   3) Back to stage 2, do looping up to the centroid value generated is permanent and members cluster does not move to cluster others.

3. Results and Discussion

3.1 Exploratory Data Analysis (EDA)

Exploratory data analysis (EDA) is needed to identify general patterns in the data. These patterns include outliers and features of the data that might be unexpected. EDA is an important first step in any data analysis.

Table 2. Summary of the Variable

	Percentage Last Education High School or Higher	Percentage Temporary Layoff Because of COVID-19	Percentage of Permanent Layoff Because of Fired	Percentage of Permanent Layoff Because of COVID-19	Minimum City Wage (IDR)
Minimum	18.55%	14.29%	3.226%	21.49%	1.558.794
First Quartile	25.71%	28.95%	9.34%	31.82%	1.672.948
Median	31.49%	41.67%	10.79%	37.40%	2.529.760
Mean	34.28%	39.67%	12.15%	37.05%	2.464.762
Third Quartile	38.59%	44.74$	14.50%	45.52%	2.683.277
Maximum	64.74%	72.73%	22.63%	51.38%	3.919.291

Table 2 shows the summary of each variable used in this research. Based on Fig. 2, it was found that the highest percentage of permanent layoff because due to firing and COVID-19 occurred in Bogor, while the lowest was in Banjar. The average percentage temporary layoff because of COVID-19 was 41.67%.

p-ISSN 2798-253X
e-ISSN 2798-3153

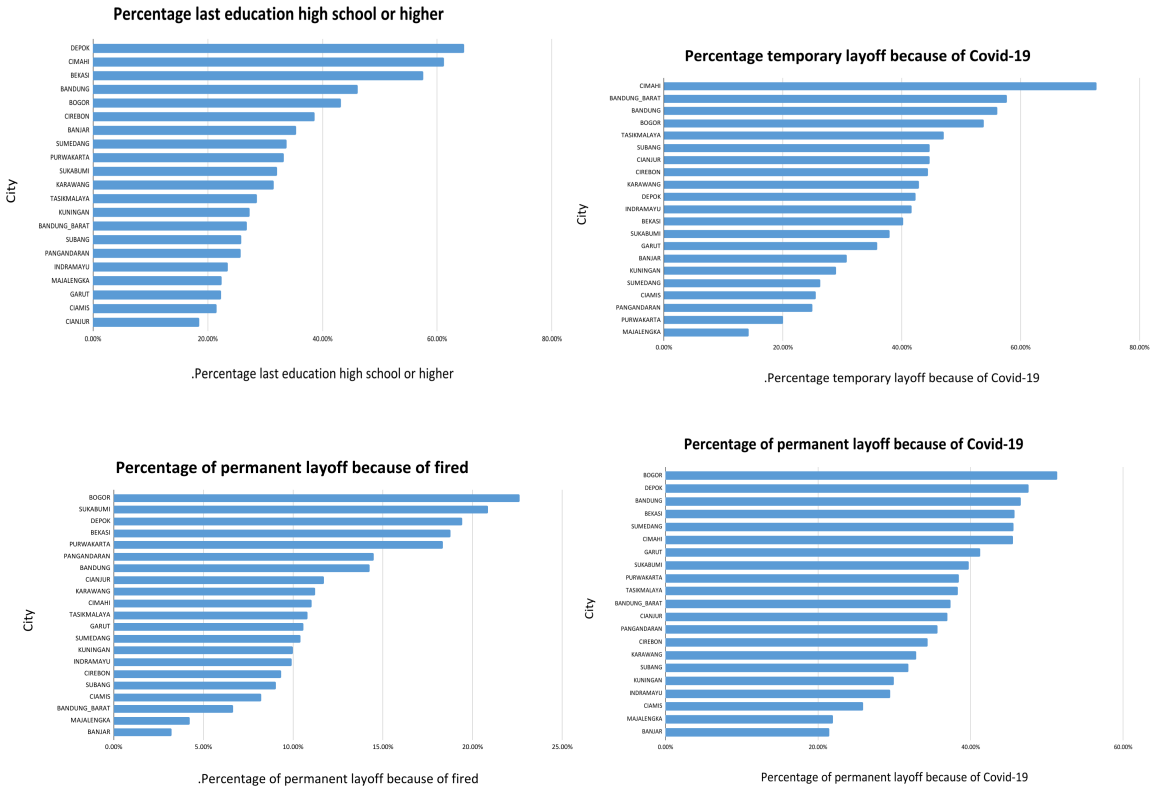

Fig. 2 Percentage of per variable for each city.

3.2 Cluster Analysis Assumption

Before conducting cluster analysis, it is necessary to fulfill two assumptions, namely the sample that can represent the population and there is no multicollinearity.

3.2.1 Representative Sample

The assumption that the sample can represent the population is obtained from the result of KMO in Table 3, while the assumption of no multicollinearity is obtained from the correlation coefficient in Table 4.

Table 3. Result of KMO Test

KMO Test	
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	0.711

The results of the KMO Test in Table 3 shows that the value is 0.711. This value ranges from 0.5 to 1, meaning that the sample can represent the population.

3.2.2 No Multicollinearity

Table 4. Multicollinearity Test

Group 1			
Dependent	**Independent**	**Tolerance**	**VIF**
Percentage last education high school or higher	Percentage temporary layoff because of COVID-19	0.589	1.698
	Percentage of permanent layoff because of fired	0.318	3.145

p-ISSN 2798-253X
e-ISSN 2798-3153

	Percentage of permanent layoff because of COVID-19	0.270	3.705
	Minimum city wage	0.525	1.905
Group 2			
Dependent	**Independent**	**Tolerance**	**VIF**
Percentage temporary layoff because of COVID-19	Percentage of permanent layoff because of fired	0.128	7.831
	Percentage of permanent layoff because of COVID-19	0.139	7.212
	Minimum city wage	0.362	2.763
	Percentage last education high school or higher	0.444	2.253
Group 3			
Dependent	**Independent**	**Tolerance**	**VIF**
Percentage of permanent layoff because of fired	Percentage of permanent layoff because of COVID-19	0.620	1.613
	Minimum city wage	0.380	2.629
	Percentage last education high school or higher	0.421	2.375
	Percentage temporary layoff because of COVID-19	0.684	1.463
Group 4			
Dependent	**Independent**	**Tolerance**	**VIF**
Percentage of permanent layoff because of COVID-19	Minimum city wage	0.353	2.835
	Percentage last education high school or higher	0.412	2.429
	Percentage temporary layoff because of COVID-19	0.680	1.470
	Percentage of permanent layoff because of fired	0.568	1.761
Group 5			
Dependent	**Independent**	**Tolerance**	**VIF**
Minimum city wage	Percentage last education high school or higher	0.701	1.427
	Percentage temporary layoff because of COVID-19	0.678	1.474
	Percentage of permanent layoff because of fired	0.133	7.510
	Percentage of permanent layoff because of COVID-19	0.135	7.422

Based on the result from Table 4, the value of VIF is less than 10 or tolerance is greater than 0.10. Therefore, it can be concluded that there is no multicollinearity between all variables. From the two assumptions of the test results above, it can be concluded that if both assumptions have been met, the sample is sufficient to represent the population, and there is no multicollinearity.

3.3 Number of Optimal Clusters

To determine the number of clusters that are optimal from those variables, the elbow method was utilized, which used the value of within sum of square (WSS) and silhouette method, as visualized in Fig. 3 and Fig. 4.

p-ISSN 2798-253X
e-ISSN 2798-3153

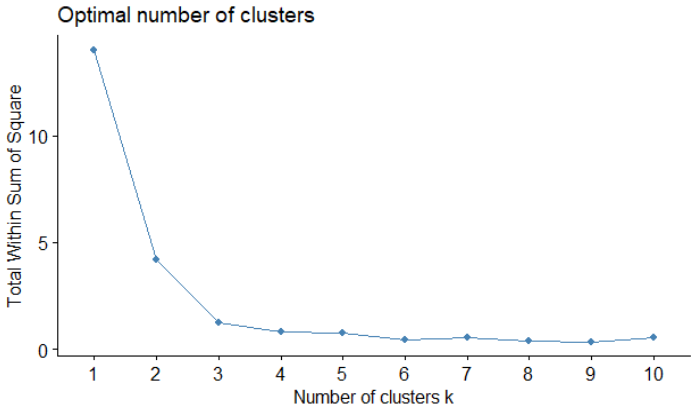

Fig. 3 Optimization number of clusters using the elbow method.

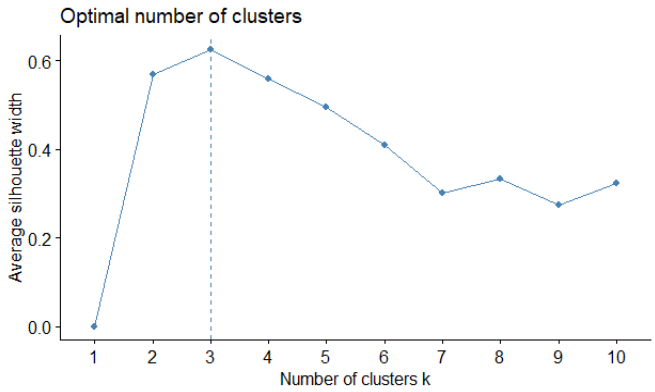

Fig. 4 Optimization number of clusters using silhouette method.

The optimization using the elbow method found that the line had a fracture at k = 3. Therefore, it can be concluded that by using WSS method, k is optimal at k = 3. Meanwhile, the optimization using the silhouette method in Fig. 4 showed that three clusters were the best option. It is because the average silhouette width with two clusters is the highest from other numbers between 1 to 10. So, for the next k-means clustering, this research used three clusters.

3.4 K-means Clustering Analysis

Based on the result of clustering using k-means clustering, the researcher obtained three clusters: cluster 1 consisted of five regencies, cluster 2 consisted of six regencies, and cluster 3 consisted of ten regencies. Each of it can be seen in Fig. 5 and Table 5 below.

Table 5. Clustering Result

Case Number	City	Cluster	Distance to Center
1	Bandung	1	0.130
2	Bandung Barat	1	0.159
3	Banjar	3	0.247
4	Bekasi	2	0.273
5	Bogor	2	0.252
6	Ciamis	3	0.190
7	Cianjur	3	0.430
8	Cimahi	1	0.338
9	Cirebon	3	0.201
10	Depok	2	0.212
11	Garut	3	0.138

Case Number	City	Cluster	Distance to Center
12	Indramayu	3	0.220
13	Karawang	2	0.316
14	Kuningan	3	0.161
15	Majalengka	3	0.248
16	Pangandaran	3	0.228
17	Purwakarta	2	0.329
18	Subang	1	0.193
19	Sukabumi	1	0.164
20	Sumedang	1	0.241
21	Tasikmalaya	3	0.222

Based on the results that have been obtained through R-studio, a plot of the visualization results of k-means clustering can be displayed.

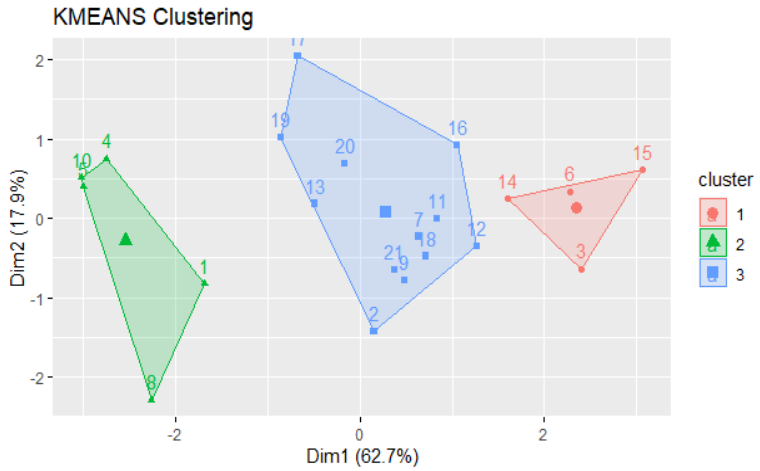

Fig. 5 Plot of the k-means clustering results.

In Fig. 5 above, the plot has three different colors that indicate the results of each cluster. The members of cluster 1 are red, the members of cluster 2 are green, and the members of cluster 3 are blue. The final results of k-means clustering using the distance between each cluster center that is shown in Table 6.

Table 6. Distance Between Final Cluster Centers

Cluster	1	2	3
1		1.041	1.928
2	1.041		0.906
3	1.928	0.906	

The percentage of temporary layoff because of COVID-19, the percentage of permanent layoff because of firing, and the percentage of permanent layoff because of COVID-19 in cluster 2 had greater averages than cluster 1 and cluster 3. Hence, the priority cities to get interventions to overcome the unemployment are Bekasi, Bogor, Depok, Karawang, and Purwakarta. The result of cluster members show in Table 7 and Fig. 6 shows Thematic map for each cluster.

Table 7. Cluster Members

Cluster	Number of City	Regency
1	6	Bandung, Bandung Barat, Cimahi, Subang, Sukabumi, Sumedang

p-ISSN 2798-253X
e-ISSN 2798-3153

Cluster	Number of City	Regency
2	5	Bekasi, Bogor, Depok, Karawang, Purwakarta
3	10	Banjar, Ciamis, Cianjur, Cirebon, Garut, Indramayu, Kuningan, Majalengka, Pangandaran, Tasikmalaya

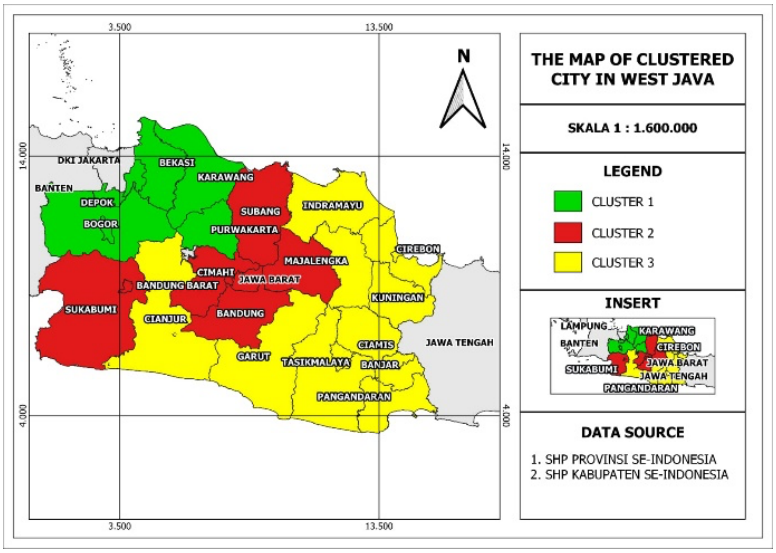

Fig. 6 Thematic map for each cluster.

Based on the clustering result, the average in each cluster can be seen in Table 8 and interpreted as follows.

a. Cluster 1 which consists of five regencies are areas with the characteristics of the lowest percentage of temporary and permanent layoff because of COVID-19, the lowest percentage of permanent layoff because of being fired, the lowest minimum city wage and the lowest percentage last education of high school or higher. Cluster 1 is grouped into cluster with low priority.
b. Cluster 2 which consists of six regencies are areas with the characteristics of the highest percentage of temporary and permanent layoff because of COVID-19, the highest percentage of permanent layoff because of being fired, the highest minimum city wage, and highest percentage last education of high school or higher. Cluster 2 is grouped into cluster with high priority.
c. Cluster 3 which consists of ten regencies are areas with the characteristics of medium percentage of temporary and permanent layoff because of COVID-19, medium percentage of permanent layoff because of fired, medium minimum city wage and medium percentage last education high school or higher. Cluster 3 is grouped into cluster with medium priority.

Table 8. Clustering Average Value

Cluster	Priority	Percentage Last Education High School or Higher	Percentage Temporary Layoff because of COVID-19	Percentage of Permanent Layoff because of fired	Percentage of Permanent Layoff because of COVID-19	Minimum City Wage (IDR)
1	Low	26.67%	24.89%	6.41%	24.81%	1,607,902

p-ISSN 2798-253X
e-ISSN 2798-3153

Cluster	Priority	Percentage Last Education High School or Higher	Percentage Temporary Layoff because of COVID-19	Percentage of Permanent Layoff because of fired	Percentage of Permanent Layoff because of COVID-19	Minimum City Wage (IDR)
2	High	54.56%	53.05%	17.22%	47.36%	3,267,962
s3	Medium	28.37%	39.02%	11.94%	36.81%	2,415,715

Increasing wage policy affects employment in individual companies also in the overall economy. Higher wage reduces employment [18]. This statement supports the fact that cities in cluster 2 have the highest average of minimum city wage, so the layoff in this pandemic era is also the highest among other clusters. Bekasi, Bogor, Depok, and Karawang are known as some of economic central in Indonesia. Hence, intervention in these cities can accelerate the improvement of the Indonesian economy.

4. Conclusion

Significant increase of open unemployment rate in West Java, as the most unemployed province in COVID-19 pandemic era, gave a negative impact to national income. To overcome the unemployment rate in West Java, some interventions must be conducted by the government. However, there are many cities which has their own characteristics. To make the intervention efficient, there are some cities that must be designed as the priority. Determining which cities were done using k-means clustering. K-means clustering algorithm is grouping data based on similar characteristics. The results showed that the open unemployment rate in West Java province in 2020 was divided into three clusters. Cluster 1 is a low priority, cluster 2 is high priority, and cluster 3 is medium priority to get interventions to overcome the unemployment. In future research, it is also necessary to make a comparison with other clustering methods to determine the best clustering model.

Acknowledgment

This research is fully supported by IPB University.

References

[1] "Menaker Ida: Jawa Barat, Provinsi Paling Banyak Pekerjanya Yang Terdampak COVID-19" BBPVP Bekasi. Accessed: June 16, 2022. [Online]. Available: https://blkbekasi.kemnaker.go.id/Berita/detail/Menaker-Ida-Jawa-Barat-Provinsi-Paling-Banyak-Pekerjanya-Yang-Terdampak-COVID-19-31ZHQ

[2] Kemnaker Data, Accessed: June. 16, 2022. [Online]. Available: https://satudata.kemnaker.go.id/data/kumpulan-data/292

[3] A.N. Ramadani, D. Sartika, and H. Herawaty, "Increase in Unemployment Rates During the COVID-19 Pandemic," *Jurnal Ilmiah Ilmu Administrasi dan Manajemen*, Vol. 15, No. 3, pp. 111–120, 2022.

[4] UNICEF, UNDP, PROSPERA, and The SMERU Research Institute, *Socioeconomic Impact of the COVID-19 Pandemic on Households in Indonesia: Three Rounds of Monitoring Surveys*. 2022. [Online]. Available: https://www.unicef.org/indonesia/media/13106/file/Socio-Economic%20Impact%20of%20COVID-19%20on%20Households%20in%20Indonesia.pdf

[5] "Produk Domestik Regional Bruto Provinsi-Provinsi di Indonesia menurtu Lapangan Usaha 2016-2020," Statistics Indonesia, 2021.

[6] K. Ishak, "Faktor-Faktor yang Mempengaruhi Pengangguran dan Inflikasinya terhadap Indeks Pembangunan di Indonesia," *Iqtishaduna*, Vol. 7, No. 1, pp. 22–38, 2018.

[7] T. Alfina, B. Santosa, and A.R. Barakbah, "Analisa Perbandingan Metode Hierarchical Clustering, K-Means, dan Gabungan Keduanya dalam Cluster Data (Studi Kasus: Problem Kerja Praktek Teknik Industri ITS)," *Jurnal Teknik ITS,* Vol. 1, No. 1, pp. 521–525, 2012.

[8] R.A. Johnson and D.W. Winchern, *Applied Multivariate Statistical Analysis*. New Jersey, USA: Prentice Hall, 2007.

[9] A.A. Matjik and I.M. Sumertajaya, *Sidik Peubah Ganda dengan Menggunakan SAS*. Bogor, Indonesia: IPB Press, 2011.

[10] C. Suhaeni, A. Kurnia, and Ristiyanti, "Perbandingan Hasil Pengelompokan Menggunakan Analisis Cluster Berhirarki, K-Means Cluster, dan Cluster Ensemble (Studi Kasus Data Indikator Pelayanan Kesehatan Ibu Hamil)," *Jurnal Media Infotama*, Vol. 14, No. 1, pp. 31–38, 2018.

[11] T. Khotimah, "Pengelompokan Surat dalam Al Quran Menggunakan Algoritma K-Means," *Jurnal Teknik Mesin Elektro dan Ilmu Komputer*, Vol. 5, No. 1, pp. 83–88, 2014.

[12] A.U. Fitriyadi, "Analisis Algoritma K-Means dan K-Medoids untuk Clustering Data Kinerja Karyawan pada Perusahaan Perumahan Nasional," *KILAT*, Vol. 10, No. 1, pp. 157–168, 2021, doi: 10.33322/kilat.v10i1.1174.

[13] F.L. Sibuea and A. Sapta, "Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering," *Jurnal Teknologi dan Sistem Informasi*, Vol. 4, No. 1, pp. 85–92, 2017, doi: 10.33330/jurteksi.v4i1.28.

[14] F. Nhita, "Comparative Study Between Parallel K-Means and Parallel K-Medoids with Message Passing Interface (MPI)," *International Journal on Information and Communication Technology*, Vol. 2, No. 2, pp. 27–36, 2016, doi: 10.21108/IJOICT.2016.22.86.

[15] R.D. Ramadhani and D.A.K. Januarita, "Evaluasi K-Means Dan K-Medoids Pada Dataset Kecil," in *Prosiding Seminar Nasional Informatika dan Aplikasinya*, Fakultas Matematika dan Ilmu Pengetahuan Universitas Jenderal Achmad Yani, 2019, 20–24.

[16] A. Widarjono, "*Analisis Statistika Multivariat Terapan*. Sleman, Indonesia: UPP STIM YKPN, 2010.

[17] Suparto, "Analisis Korelasi Variabel-Variabel yang Mempengaruhi Siswa dalam Memilih Perguruan Tinggi," in *Seminar Nasional Sains dan Teknologi Terapan II*, Institut Teknologi Adhi Tama Surabaya, 2014, 469–478.

[18] D.S. Hamermesh. "Do Labor Costs Affect Companies' Demand for Labor?" IZA World of Labor. Accessed: June. 18, 2022 [Online]. Available: https://wol.iza.org/articles/do-labor-costs-affect-companies-demand-for-labor/long

p-ISSN 2798-253X
e-ISSN 2798-3153