Analisis K-Means Clustering untuk Klasterisasi TPT Kabupaten/Kota di Jawa Barat 2021–2023

Ainani Tajriyan Muntaharridwan *, Rafly Isman Putra, Husna Khairiyah Az Zahra, Aulia Athhar Zakiyyan, Renita Maharani

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

ainani.tajriyan80@gmail.com

Abstract. The Open Unemployment Rate (TPT) is a crucial indicator reflecting the labor market and economic conditions of a region. West Java Province has consistently recorded relatively high TPT figures over the past three years, despite showing a downward trend. This study aims to cluster districts/cities in West Java based on their TPT data from 2021 to 2023 using the K-Means Clustering algorithm. The data used is secondary data obtained from the Central Statistics Agency (BPS). The analytical process includes assumption testing, data normalization using Z-Score, and determining the optimal number of clusters through the Silhouette Coefficient method. The results show that two optimal clusters were formed: the first cluster consists of five regions with relatively low TPT, while the second cluster includes twenty-two regions with relatively high TPT. The clustering visualization demonstrates a clear separation between clusters, and silhouette values above 0.5 indicate strong clustering performance. This research provides a data-driven basis for local governments to formulate more targeted and effective employment policies. It also contributes to future studies in applying data mining techniques to social and economic issues.

Keywords: *K-Means Clustering, Open Unemployment, TPT West Java.*

Abstrak. Tingkat Pengangguran Terbuka (TPT) merupakan indikator penting yang mencerminkan kondisi ketenagakerjaan dan perekonomian suatu wilayah. Provinsi Jawa Barat tercatat memiliki angka TPT yang cukup tinggi dalam tiga tahun terakhir, meskipun menunjukkan tren penurunan. Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Jawa Barat berdasarkan kesamaan tingkat pengangguran terbuka pada tahun 2021 hingga 2023 menggunakan algoritma K-Means Clustering. Data yang digunakan merupakan data sekunder dari Badan Pusat Statistik (BPS). Proses analisis diawali dengan uji asumsi, normalisasi data menggunakan Z-Score, serta validasi jumlah cluster optimal menggunakan metode Silhouette Coefficient. Hasil penelitian menunjukkan bahwa dua cluster optimal terbentuk: cluster pertama terdiri dari lima kabupaten/kota dengan TPT relatif rendah, sedangkan cluster kedua terdiri dari dua puluh dua wilayah dengan TPT relatif tinggi. Visualisasi hasil clustering menunjukkan pemisahan yang jelas antar cluster, dan validasi dengan nilai silhouette rata-rata > 0,5 menunjukkan kualitas pengelompokan yang baik. Penelitian ini diharapkan menjadi dasar rekomendasi bagi pemerintah daerah dalam merancang kebijakan ketenagakerjaan yang lebih terarah dan efisien, serta sebagai referensi dalam pengembangan penelitian lanjutan menggunakan metode data mining dalam bidang sosial ekonomi.

Kata Kunci: *K-Means Clustering, Pengangguran Terbuka, TPT Jawa Barat*.

*ainani.tajriyan80@gmail.com

A. Pendahuluan

Tingkat Pengangguran Terbuka (TPT) merupakan indikator makroekonomi penting yang merefleksikan kapasitas pasar tenaga kerja dalam menyerap angkatan kerja di suatu wilayah. Semakin rendah nilai TPT, semakin baik pula kinerja sektor ekonomi dan produktivitas wilayah tersebut. Sebaliknya, tingginya angka pengangguran menjadi sinyal adanya ketidakseimbangan antara penawaran dan permintaan tenaga kerja, serta dapat mengindikasikan kurang efektifnya program-program ketenagakerjaan dan ekonomi yang dijalankan (Sukirno, 1994).

Provinsi Jawa Barat sebagai salah satu provinsi dengan jumlah penduduk terbesar di Indonesia menghadapi tantangan besar dalam hal pengangguran. Menurut data Badan Pusat Statistik (BPS) pada Agustus 2023, TPT Jawa Barat mencapai 7,44%, menurun dari 8,31% pada 2022 dan 9,82% pada 2021. Walaupun penurunan ini menggambarkan adanya perbaikan, namun angka tersebut masih termasuk tinggi dibandingkan dengan rata-rata nasional, serta menunjukkan adanya disparitas antar wilayah kabupaten/kota di dalam provinsi (Badan Pusat Statistik, 2023).

Analisis terhadap TPT tidak cukup hanya dilakukan dengan melihat angka agregat, melainkan perlu juga dilakukan pengelompokan berdasarkan karakteristik kemiripan antar daerah. Untuk tujuan tersebut, metode unsupervised learning seperti K-Means Clustering dapat menjadi alat bantu yang sangat berguna (Han et al., 2012). K-Means dapat mengelompokkan daerah-daerah ke dalam beberapa cluster yang homogen berdasarkan nilai TPT selama tiga tahun terakhir. Dengan hasil pengelompokan tersebut, pemerintah provinsi dan kabupaten/kota dapat lebih fokus dalam menyusun program penurunan pengangguran berdasarkan kelompok daerah dengan karakteristik serupa.

Penelitian ini bertujuan untuk mengaplikasikan metode K-Means Clustering dalam mengelompokkan 27 kabupaten/kota di Jawa Barat berdasarkan data TPT tahun 2021–2023. Penelitian ini diharapkan dapat memberikan gambaran empiris mengenai kondisi pengangguran di wilayah tersebut serta memberikan rekomendasi kebijakan yang lebih tepat sasaran.

B. Metode

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode analisis klaster berbasis algoritma K-Means (Jain, 2010). Langkah-langkah utama dalam penelitian ini meliputi:

1. **Pengumpulan Data**
   Data TPT tahun 2021–2023 dikumpulkan dari situs resmi BPS Jawa Barat (https://jabar.bps.go.id). Dataset mencakup seluruh 27 kabupaten/kota di provinsi tersebut. Data diklasifikasikan sebagai data sekunder.
2. **Uji Asumsi**
   Uji asumsi dilakukan dengan metode Kaiser Meyer Olkin (KMO). Nilai KMO sebesar 0,776 menunjukkan bahwa data memenuhi syarat keandalan untuk dilakukan analisis lebih lanjut (Hair et al., 2010).
3. **Pemrosesan dan Normalisasi Data**
   Pemrosesan data meliputi:
   a. Deteksi pencilan menggunakan boxplot (Witten et al., 2011).
   b. Pengecekan nilai hilang (missing value), hasilnya adalah nihil.
   c. Normalisasi data menggunakan Z-Score normalization untuk menyamakan skala antar tahun (Tan et al., 2006).

$$Z = \frac{X-\mu}{\sigma} \quad \text{... (1)}$$

   Keterangan:
   X = Nilai Asli
   μ = Rata-rata
   σ = Simpangan Baku

4. **Penentuan Jumlah Cluster**
   Jumlah cluster ditentukan dengan metode Silhouette Coefficient. Grafik hasil menunjukkan nilai maksimum pada k=2, yang berarti dua cluster adalah jumlah optimal (Nahdliyah et al., 2022).

5. Analisis Cluster

Manual: dengan perhitungan jarak Euclidean antar titik dan iterasi hingga konvergen (MacQueen, 1967).

$$d(ij) = \sqrt{\sum_{k=1}^{n} \left(x_{ik} - x_{jk}\right)^2} \quad \ldots (2)$$

Keterangan:

x_{ik} dan x_{jk} adalah komponen ke-k dari titik i dan j.

Menggunakan R-Studio: memanfaatkan library tidyverse, cluster, dan factoextra untuk menghasilkan visualisasi dan evaluasi kualitas cluster (Sani, 2018).

6. Visualisasi dan Validasi

Visualisasi hasil cluster dilakukan menggunakan fviz_cluster. Validasi hasil dilakukan dengan mengevaluasi nilai Silhouette score. Nilai di atas 0.5 dianggap memadai.

$$s(i) = \frac{b(i) - a(i)}{\max\left(a(i), b(i)\right)} \quad \ldots (3)$$

Keterangan:

a(i) : jarak rata-rata i ke semua titik dalam cluster yang sama

b(i) : jarak rata-rata i ke titik dalam cluster terdekat lainnya.

C. Hasil Penelitian dan Pembahasan

Bagian Uji Asumsi K-Means Clustering

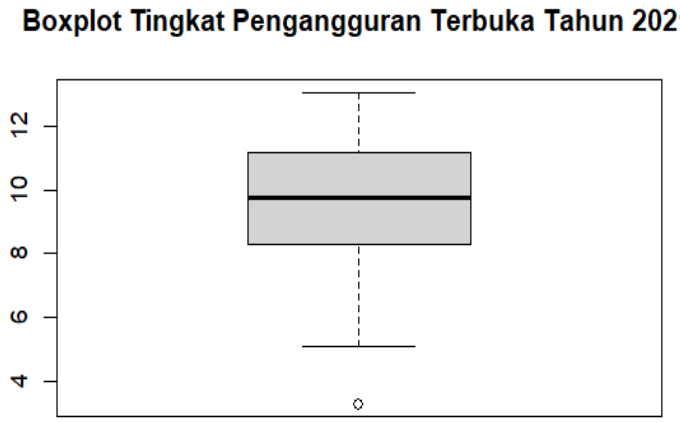
Tabel 1. Hasil Pengujian KMO

	Koefisien	***p-value***
KMO	0.776	0.000

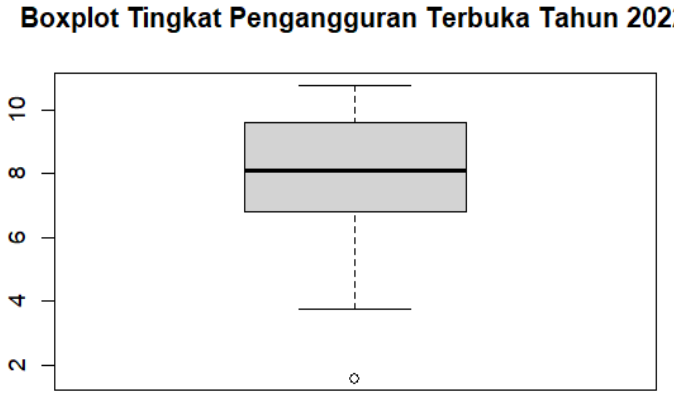
Berdasarkan tabel 1 di atas, diperoleh koefisien KMO sebesar 0.776 yang berkisar diantara 0.5-1, artinya asumsi sampel representatif terpenuhi atau sampel telah mewakili populasi yang ada. Dalam penelitian ini, peneliti hanya menggunakan satu variabel yaitu tingkat pengangguran terbuka di Jawa Barat yang terdiri dari data tahun 2021, 2022, dan 2023. Jelas sekali bahwa data tersebut tidak dapat dilakukan pengujian multikolinearitas karena hanya terdapat satu variabel. Kemudian data tahun 2021, 2022, dan 2023 merupakan data time-series yang dapat diimplementasikannya melalui pengujian autokorelasi. Dalam analisis cluster tidak diperlukan pengujian autokorelasi sehingga diasumsikan bahwa nonmultikolinieritas dalam kasus ini sudah terpenuhi.

Pemrosesan dan Normalisasi Data

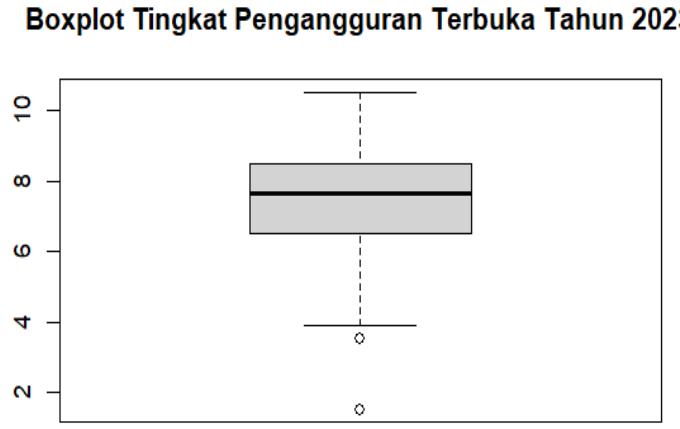
Fitur atau atribut yang digunakan dalam penelitian ini yaitu data kabupaten/kota di Jawa Barat dan data tingkat pengangguran terbuka di Jawa Barat tahun 2021-2023. Pendeteksian pencilan dilakukan dengan menggunakan boxplot sebagai berikut:



Gambar 2. Boxplot Tingkat Pengangguran Terbuka Tahun 2021



Gambar 3. Boxplot Tingkat Pengangguran Terbuka Tahun 2022



Gambar 4. Boxplot Tingkat Pengangguran Terbuka Tahun 2023

Berdasarkan pendeteksian pencilan menggunakan boxplot di atas, diketahui bahwa terdapat 1 kabupaten/kota yang merupakan pencilan di tahun 2021 dan 2022 yaitu Kabupaten Pangandaran, serta terdapat 2 kabupaten/kota yang merupakan pencilan di tahun 2023 yaitu Kabupaten Pangandaran dan Kabupaten Ciamis. Pada kasus ini data pencilan tetap diikutsertakan dalam analisis selanjutnya sebab menggunakan fitur atau atribut wilayah berupa kabupaten/kota yang tidak mungkin untuk dihilangkan.

Tabel 2. Hasil Missing Value

	Jumlah
***Missing Value* (NA)**	0

Berdasarkan hasil pendeteksian data hilang atau missing value di atas, diperoleh bahwa tidak ada data yang hilang dari dataset yang akan dianalisis. Sehingga tidak perlu dilakukan penanganan terhadap data yang hilang atau missing value. Normalisasi data dilakukan menggunakan metode normalisasi skor Z standar atau Z-score normalization (Standardization). Adapun hasil dari normalisasi data adalah sebagai berikut:

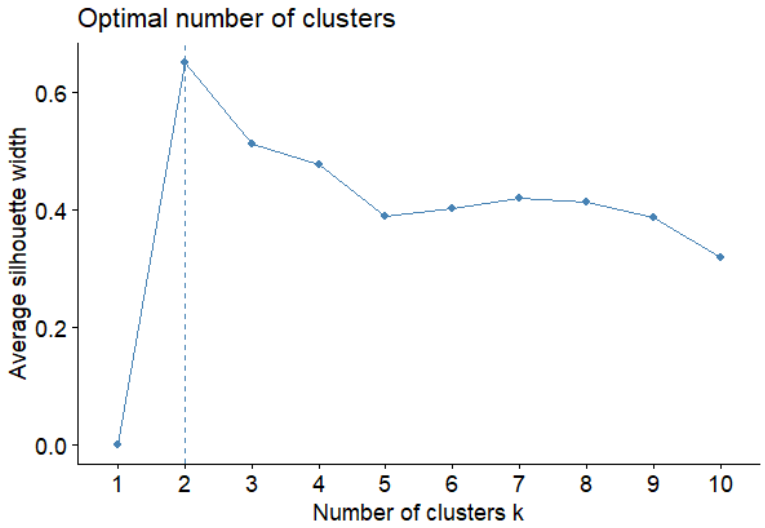
Tabel 3. Hasil Normalisasi Data

Kabupaten/Kota	**Normalisasi Skor Z**		
	2021	**2022**	**2023**
Bogor	1.1582	1.2174	0.6348
Sukabumi	0.0447	-0.0133	0.0663
Cianjur	-0.0333	0.2611	0.2591
Bandung	-0.4442	-0.3521	-0.3292
Garut	-0.2963	-0.0862	0.0712
Tasikmalaya	-1.3316	-1.5571	-1.6294
Ciamis	-1.7836	-1.7372	-1.8124
Kuningan	0.9363	0.8615	1.1391
Cirebon	0.4022	0.1325	0.2294
Majalengka	-1.5165	-1.5614	-1.5157
Sumedang	-0.0908	-0.0348	-0.1216
Indramayu	-0.4524	-0.5622	-0.3589
Subang	0.1516	-0.0133	0.2294
Purwakarta	0.5337	0.4069	0.2640
Karawang	0.9979	0.8872	0.8721
Bekasi	0.2830	1.0759	0.8326
Bandung Barat	0.9240	0.7843	0.4568
Pangandaran	-2.5272	-2.6763	-2.8011
Kota Bogor	0.9815	1.2774	1.0897
Kota Sukabumi	0.5665	0.4412	0.6645
Kota Bandung	0.8459	0.7500	0.8128

Kabupaten/Kota	Normalisasi Skor Z		
	2021	2022	2023
Kota Cirebon	0.4638	0.2654	0.2344
Kota Bekasi	0.6076	0.4326	0.3530
Kota Depok	0.1475	0.0081	-0.1067
Kota Cimahi	1.5074	1.2731	1.6483
Kota Tasikmalaya	-0.7153	-0.5065	-0.3144
Kota Banjar	-1.3604	-0.9739	-0.8681

Analisis K-Means Clustering

Penentuan jumlah cluster dilakukan menggunakan Silhouette Coefficient dimana k yang dipilih merupakan k dengan skor silhouette tertinggi.



Gambar 5. Grafik Silhouette

Berdasarkan grafik di atas, terlihat bahwa nilai k dengan skor silhouette tertinggi adalah ketika k = 2. Artinya jumlah cluster yang paling optimal untuk kasus ini adalah 2 cluster. Sehingga hasil pembentukan cluster terbaik berdasarkan data tingkat pengangguran terbuka di Jawa Barat tahun 2021-2023 yaitu sebanyak 2 cluster. Berdasarkan pengelompokan menggunakan perhitungan manual dan software R-Studio, diperoleh bahwa cluster 1 terdiri dari 5 kabupaten/kota, sedangkan cluster 2 terdiri dari 22 kabupaten/kota. Cluster dikategorikan berdasarkan tingkat pengangguran terbuka rendah dan tinggi. Adapun kategori cluster dapat dilihat berdasarkan nilai rata-rata tingkat pengangguran terbuka dari setiap cluster-nya.

Tabel 4. Rata-Rata Cluster 1

***Cluster* 1**	
Kabupaten/Kota	**TPT**
Tasikmalaya	4.74
Ciamis	4.11
Majalengka	4.6633
Pangandaran	2.1100
Kota Banjar	5.6833
Rata-rata	**4.2613**

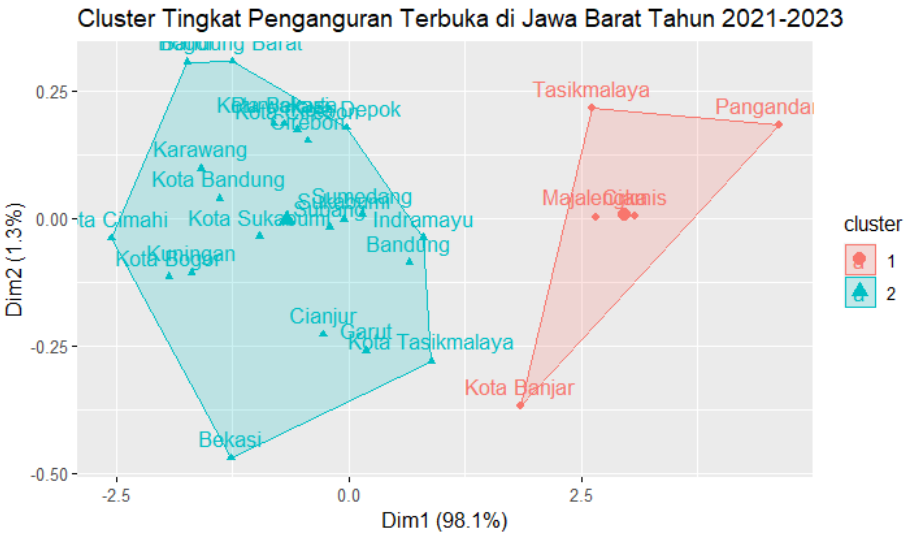
Tabel 5. Rata-Rata Cluster 2

***Cluster* 2**	
Kabupaten/Kota	**TPT**
Bogor	10.4433
Sukabumi	8.2000
Cianjur	8.4800
Bandung	7.2733
Garut	7.8700
Kuningan	10.3267
Cirebon	8.7133
Sumedang	7.9467
Indramayu	7.0833
Subang	8.3967
Purwakarta	9.0567
Karawang	10.2167
Bekasi	9.7567
Bandung Barat	9.7967
Kota Bogor	10.6533
Kota Sukabumi	9.3800
Kota Bandung	9.9467
Kota Cirebon	8.8700

Cluster 2	
Kabupaten/Kota	**TPT**
Kota Bekasi	9.1967
Kota Depok	8.1833
Kota Cimahi	11.4533
Kota Tasikmalaya	6.9433
Rata-rata	**9.0085**

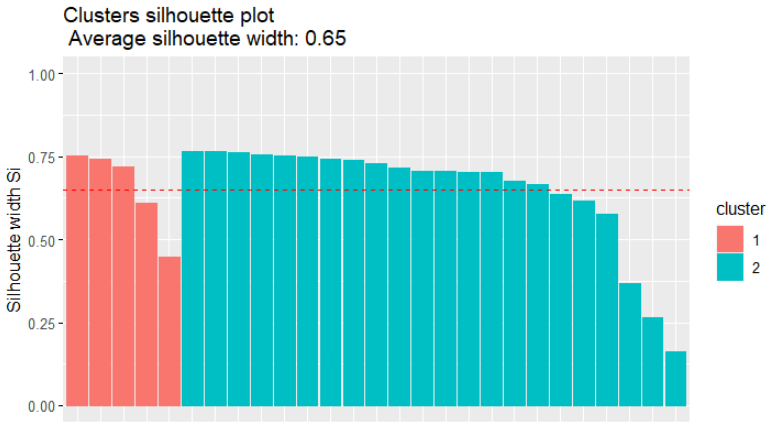
Berdasarkan hasil analisis cluster menggunakan metode K-Means clustering di atas diperoleh rata-rata tingkat pengangguran terbuka di cluster 1 adalah sebesar 4.26%, sedangkan rata-rata tingkat pengangguran terbuka di cluster 2 adalah sebesar 9.01%. Sehingga dapat disimpulkan bahwa cluster 1 merupakan cluster dengan tingkat pengangguran terbuka kategori rendah dan cluster 2 merupakan cluster dengan tingkat pengangguran terbuka kategori tinggi.

Visualisasi Cluster



Gambar 6. Visualisasi Cluster Tingkat Pengangguran Terbuka di Jawa Barat

Berdasarkan visualisasi cluster di atas, diketahui bahwa cluster 1 terdiri dari 5 kabupaten/kota yaitu Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Majalengka, Kabupaten Pangandara, dan Kota Banjar. Kemudian cluster 2 terdiri dari 22 kabupaten/kota yaitu Kabupaten Bogor, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Bandung, Kabupaten Garut, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kabupaten Bandung Barat, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, dan Kota Tasikmalaya.



Gambar 7. Validasi Cluster Tingkat Pengangguran Terbuka Provinsi Jawa Barat

Plot siluet cluster di atas menunjukkan bukti bahwa penggunaan 2 cluster menghasilkan pengelompokan yang baik karena tidak ada lebar siluet yang negatif dan sebagian besar bernilai lebih besar dari 0.5.

D. Kesimpulan

Penelitian ini menyimpulkan bahwa metode K-Means clustering efektif digunakan untuk mengelompokkan kabupaten/kota di Provinsi Jawa Barat berdasarkan data Tingkat Pengangguran Terbuka (TPT) tahun 2021–2023. Proses dimulai dengan persiapan data, termasuk pemilihan fitur, deteksi pencilan dan nilai hilang, serta normalisasi data menggunakan Z-score. Jumlah cluster optimal ditentukan menggunakan Silhouette Coefficient, yang menghasilkan dua cluster. Cluster pertama terdiri dari lima wilayah dengan TPT rendah: Kabupaten Tasikmalaya, Ciamis, Majalengka, Pangandaran, dan Kota Banjar. Sementara itu, cluster kedua berisi 22 kabupaten/kota dengan TPT tinggi seperti Kabupaten Bogor, Bandung, dan Kota Depok. Temuan ini menunjukkan bahwa sebagian besar wilayah di Jawa Barat masih berada dalam kategori TPT tinggi, sehingga hasil pengelompokan ini dapat digunakan sebagai dasar bagi Pemerintah Provinsi Jawa Barat dalam menyusun kebijakan ketenagakerjaan yang lebih terarah. Mengingat keterbatasan penelitian, hasil ini sebaiknya digunakan sebagai bahan evaluasi awal. Diharapkan pemerintah melakukan tindak lanjut melalui kebijakan solutif yang disesuaikan dengan karakteristik daerah serta mendorong kajian lanjutan terhadap faktor-faktor penyebab tingginya pengangguran di wilayah tersebut.

Daftar Pustaka

Camartya, D., & Achmad, A. I. (2022). Analisis Korespondensi pada Jumlah Pengangguran Terbuka Menurut Kabupaten/Kota Berdasarkan Pendidikan Tertinggi. *Jurnal Riset Statistika*, 119–128. https://doi.org/10.29313/jrs.v2i2.1424

Khofifah, H. N. (2022). Robust Spatial Durbin Model (RSDM) untuk Pemodelan Tingkat Pengangguran Terbuka (TPT) di Provinsi Jawa Barat. *Jurnal Riset Statistika*, *1*(2), 135–142. https://doi.org/10.29313/jrs.v1i2.522

Badan Pusat Statistik. (2023). Tingkat Pengangguran Terbuka Jawa Barat. https://jabar.bps.go.id

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2010). Multivariate Data Analysis (7th ed.). Pearson Education.

Mulyadi. (2003). Ekonomi Sumber Daya Manusia dalam Perspektif Pembangunan. Jakarta: Rajagrafindo Persada.

Nahdliyah, M. A., Widiharih, T., & Prahutama, A. (2022). Metode K-Medoids Clustering dengan Validasi Silhouette Index dan C-Index (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018). Semarang: Departemen Statistika FSM UNDIP.

Noviyanto. (2020). Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia. Paradigma: Jurnal Komputer dan Informatika, 21(2).

Sani, A. (2018). Penerapan Metode K-Means Clustering pada Perusahaan. JUTI: Jurnal Ilmiah Teknologi Informasi, 16(2), 45–50.

Sukirno, S. (1994). Pengantar Teori Ekonomi Makro. Jakarta: Raja Grafindo Persada.

Dharshinni, D., et al. (2023). Penerapan Algoritma K-Means pada Data Pengangguran di Jawa Barat. DSI: Jurnal Data Science Indonesia.

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651–666.

Lloyd, S. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, 281–297.

Rokach, L., & Maimon, O. (2005). Clustering Methods. In Data Mining and Knowledge Discovery Handbook (pp. 321–352). Springer.

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). Introduction to Data Mining. Pearson Addison Wesley.

Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining: Practical Machine Learning Tools and Techniques (3rd ed.). Morgan Kaufmann

Viqri, Z. S. C., & Kurniati, E. (2023). Perbandingan Penerapan Metode Fuzzy Time Series Model Chen-Hsu dan Model Lee dalam Memprediksi Kurs Rupiah terhadap Dolar Amerika. *DataMath: Journal of Statistics and Mathematics*, *1*(1), 19–26. https://doi.org/10.29313/datamath.v1i1.12