

Pengelompokan Kabupaten/Kota di Provinsi Jawa Barat Berdasarkan Tingkat Pengangguran Terbuka Tahun 2023 Menggunakan *K-Means Clustering*

Gusmiati Husnul Fatimah¹, Glagah Eskacakra Setyowisnu²

^{1,2} Program Studi Matematika, Universitas Jenderal Soedirman

Jl. Profesor DR. HR Boenyamin No. 708, Grendeng, Purwokerto Utara, Banyumas, 53122

Email: gusmiati.fatimah@mhs.unsoed.ac.id¹, glagah.setyowisnu@unsoed.ac.id²

Korespondensi penulis: glagah.setyowisnu@unsoed.ac.id

Submitted : 31 Januari 2025

Accepted : 11 Agustus 2025

Published : 11 Agustus 2025

Abstrak

Pengangguran adalah salah satu masalah serius yang dihadapi oleh banyak negara, termasuk Indonesia. Dari banyak provinsi di Indonesia, Jawa Barat termasuk dalam provinsi dengan jumlah penduduk yang besar dengan variasi pengangguran yang tinggi. Mengingat masalah tersebut cukup besar, pemerintah perlu merancang kebijakan efektif, spesifik, dan tepat sasaran dengan acuan yang sesuai. Oleh karena itu, pada penelitian ini dilakukan analisis pengelompokan kabupaten/kota di Provinsi Jawa Barat berdasarkan pengangguran terbuka dengan menggunakan metode *K-means clustering*. Data yang digunakan mencakup jumlah pengangguran terbuka usia 15 tahun ke atas menurut tingkat pendidikan yang diperoleh dari Badan Pusat Statistik (BPS) tahun 2023. Analisis dilakukan dengan bantuan software IBM SPSS Statistic 26 dan RStudio. Penentuan banyak kluster optimal dilakukan menggunakan metode *elbow*, yang dilakukan dengan menghitung nilai *Within-cluster Sum of Squares*, kemudian hasilnya divisualisasikan dalam bentuk kurva untuk membantu menentukan banyak kluster optimal. Perhitungan tersebut menghasilkan tiga kluster dengan karakteristik pengangguran yang berbeda, di mana kluster 1, 2, dan 3 secara berturut-turut merupakan kluster pengangguran rendah, tinggi, dan sedang, serta terdiri dari 16, 7, dan 3 kabupaten/kota pada setiap kluster tersebut. Hasil ini diharapkan dapat menjadi acuan pemerintah daerah dalam merancang kebijakan untuk menurunkan tingkat pengangguran di Jawa Barat.

Kata Kunci: *K-means clustering*, pengangguran terbuka, metode *elbow*, tingkat pendidikan.

Abstract

Unemployment is one of the serious problems faced by many countries, including Indonesia. West Java is one of the many provinces in Indonesia which has large population with high unemployment variations. Given the large size of the problem, the government needs to design effective, specific, and targeted policies with appropriate references. Therefore, this study

conducted an analysis of the grouping of districts/cities in West Java Province based on open unemployment using the method of K-means clustering. The data used includes the number of open unemployment aged 15 years and above based on educational level obtained from Badan Pusat Statistik (BPS) in 2023. The analysis part was conducted with the help of two applications named IBM SPSS Statistic 26 and RStudio software. The optimal number of clusters was determined using the elbow method, which is done by calculating the Within-cluster Sum of Squares value and then visualizing in the form of a curve to help determine the optimal number of clusters. The calculation produces three clusters with different unemployment characteristics, where the cluster 1, 2, and 3 are respectively low, high, and medium unemployment clusters, and consist of 16, 7, and 3 districts/cities in each cluster. These results are expected to be a reference for local governments in designing policies to reduce the unemployment rate in West Java.

Keywords: K-means clustering, open unemployment, elbow method, educational level.

1. Pendahuluan

Pengangguran menjadi salah satu masalah besar yang dihadapi oleh banyak negara berkembang, termasuk Indonesia [1]. Salah satu jenis pengangguran berdasarkan sifatnya adalah pengangguran terbuka yang mana merupakan kondisi ketika seseorang dalam usia angkatan kerja tidak memiliki pekerjaan sama sekali, meskipun sedang aktif mencari pekerjaan atau siap untuk bekerja [2]. Pengangguran terbuka tidak hanya mempengaruhi kesejahteraan individu, tetapi juga berdampak pada stabilitas sosial dan ekonomi secara keseluruhan. Salah satu faktor yang sering dikaitkan dengan tingkat pengangguran terbuka adalah tingkat pendidikan terakhir yang ditamatkan. Umumnya, individu dengan latar belakang pendidikan yang lebih tinggi memiliki peluang lebih besar untuk mendapatkan pekerjaan. Namun, di beberapa daerah tingkat pengangguran tetap tinggi meskipun tingkat pendidikan cukup baik. Hal ini menunjukkan bahwa ada faktor lain yang berpengaruh selain faktor pendidikan, seperti pertumbuhan penduduk, kondisi ekonomi daerah, ketersediaan lapangan kerja, dan keahlian yang dimiliki oleh tenaga kerja [3].

Dari sekian banyak provinsi di Indonesia, Jawa Barat merupakan salah satu provinsi dengan jumlah penduduk terbanyak di Indonesia dengan variasi tingkat pengangguran terbuka antar kabupaten/kota yang cukup signifikan. Beberapa daerah memiliki tingkat pengangguran yang relatif rendah seperti Kota Banjar, sementara daerah lain menghadapi masalah pengangguran yang cukup serius seperti Kabupaten Bogor [4]. Perbedaan ini menunjukkan adanya ketimpangan ekonomi dan akses terhadap peluang kerja di Provinsi Jawa Barat. Analisis pengelompokan kabupaten/kota berdasarkan tingkat pengangguran terbuka dapat membantu dalam mengidentifikasi daerah-daerah yang memerlukan perhatian khusus dan penyusunan kebijakan yang lebih tepat sasaran.

Beberapa penelitian terkait pengelompokan telah dilakukan menggunakan beberapa metode. Dari beberapa metode yang telah tersaji, salah satu metode yang telah digunakan adalah *K-means clustering*, yang merupakan metode pengelompokan data dengan tujuan untuk membagi data dalam beberapa kelompok (klaster) berdasarkan karakteristik tertentu. Penelitian terkait algoritma *K-means* telah dilakukan oleh [5], yang menggunakan metode tersebut untuk mengelompokkan kabupaten/kota di Provinsi Jawa Tengah berdasarkan banyak kasus COVID-19. Penelitian ini menghasilkan klaster-klaster yang membantu pemerintah dalam mengidentifikasi daerah yang perlu perhatian lebih karena tingginya kasus aktif dan kematian. Penelitian serupa oleh [6] pada wilayah Provinsi Banten mengelompokkan daerah berdasarkan indikator Indeks Pembangunan

Manusia (IPM), yang memungkinkan pengaturan kebijakan pembangunan lebih tepat sasaran berdasarkan kesamaan karakteristik IPM. Penelitian selanjutnya oleh [3] menggunakan metode yang sama untuk menganalisis tingkat pengangguran di Provinsi Jawa Barat pada tahun 2017–2022 berdasarkan tingkat pendidikan. Hasil dari penelitian ini berupa klasifikasi pengangguran rendah, sedang, dan tinggi yang diharapkan dapat menjadi masukan bagi pemerintah dalam menentukan strategi pengurangan pengangguran di wilayah tersebut. Lebih lanjut, penerapan algoritma *K-means* pada dataset pengangguran terbuka berdasarkan pendidikan di Provinsi Jawa Barat juga telah dilakukan pada penelitian [7], yang menunjukkan 3 (tiga) pembagian tingkat pengangguran menggunakan dataset tahun 2019-2022.

Berdasarkan uraian di atas, maka tujuan dari penelitian ini adalah menerapkan metode *K-means clustering* sebagai metode analisis yang efektif untuk mengelompokkan kabupaten/kota di Provinsi Jawa Barat berdasarkan kesamaan karakteristik pada pengangguran terbuka. Berbeda dengan [3], penelitian ini menggunakan data pengangguran terbuka Provinsi Jawa Barat berdasarkan tingkat pendidikan tahun 2023. Pengelompokan akan dilakukan dengan penyajian yang cukup detail. Kemudian, akan dilakukan beberapa kajian berdasarkan hasil penelitian yang telah didapatkan. Hasil dari pengelompokan ini dapat dimanfaatkan dalam merancang strategi penanggulangan pengangguran yang lebih terfokus dan spesifik untuk setiap kelompok. Lebih lanjut, pengelompokan yang diperoleh diharapkan dapat memberikan gambaran yang lebih jelas mengenai kondisi pengangguran berdasarkan tingkat pendidikan di Jawa Barat. Dengan demikian, pemerintah daerah dapat terbantu dalam perancangan kebijakan yang lebih efektif dan tepat sasaran untuk mengurangi tingkat pengangguran.

2. Metode Penelitian

2.1 Data

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa jumlah penduduk usia 15 tahun ke atas yang termasuk pengangguran terbuka menurut kabupaten/kota dan pendidikan tertinggi yang ditamatkan di Jawa Barat tahun 2023. Pendidikan tertinggi yang ditamatkan meliputi sekolah dasar ke bawah, sekolah menengah pertama, sekolah menengah atas/ sederajat, dan diploma I/II/III/akademi/universitas. Data tersebut diperoleh dari publikasi Badan Pusat Statistik Provinsi Jawa Barat.

2.2 Analisis Data

Analisis data dilakukan dengan bantuan *software* IBM SPSS *Statistic* 26 dan *Rstudio*. Langkah-langkah yang dilakukan untuk mengelompokkan kabupaten/kota di Jawa Barat berdasarkan data tingkat pengangguran terbuka menggunakan metode *K-means clustering* mencakup: (1) menentukan statistik deskriptif; (2) melakukan standarisasi data; (3) melakukan uji statistik Hopkins; (4) melakukan uji asumsi klasik analisis klaster; (5) menentukan banyak klaster k yang paling optimal; (6) menentukan pusat klaster (*centroid*) awal secara acak; (7) melakukan proses iterasi; (8) menginterpretasi klaster dan karakteristiknya; dan (9) melakukan validasi akhir terhadap hasil klaster menggunakan metode *Silhouette Coefficient*.

2.3 Uji Statistik Hopkins

Uji statistik Hopkins diperlukan untuk memastikan bahwa pengelompokan data yang ada bukan sekedar data acak, namun memiliki makna. Uji ini dapat mengukur seberapa besar kemungkinan dataset akan berdistribusi secara seragam [8]. Untuk memastikannya, dilakukan pengujian hipotesis untuk melihat pola distribusi data dan signifikansi dari klaster yang terbentuk. Hipotesis nol menyatakan bahwa dataset asli terdistribusi secara seragam, yang berarti tidak terdapat klaster yang bermakna. Sebaliknya, hipotesis alternatif menyatakan bahwa dataset tidak terdistribusi secara seragam, yang berarti terdapat klaster yang bermakna. Apabila nilai statistik Hopkins mendekati 1, maka hipotesis nol dapat ditolak yang artinya terdapat kemampuan klaster yang bermakna [9]. Rumus untuk menghitung nilai statistik Hopkins H adalah sebagai berikut

$$H = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i}, \quad (1)$$

dengan n , x_i , dan y_i berturut-turut menyatakan jarak yang diambil secara acak, jarak dari titik observasi asli ke titik observasi terdekat lainnya dalam dataset, dan jarak dari titik acak yang dihasilkan secara seragam ke titik observasi terdekat dalam dataset.

2.4 Metode *Elbow*

Metode *elbow* merupakan salah satu metode untuk menentukan banyak klaster optimal pada *K-means clustering*. Prinsip dari metode ini adalah memulai dengan menetapkan $k = 2$ sebagai banyak klaster awal optimal. Selanjutnya, terus tingkatan nilai sampai mencapai batas maksimal yang telah ditentukan untuk memperkirakan banyak klaster optimal [9]. Proses metode *elbow* dilakukan dengan menghitung *Within-cluster Sum of Squares* (WCSS) untuk berbagai banyak klaster. Kemudian, WCSS divisualisasikan dalam bentuk kurva yang menunjukkan hubungan antara banyak klaster dan WCSS. Kurva ini menunjukkan penurunan yang signifikan di awal, kemudian penurunannya mulai melambat. Titik di mana laju penurunan mulai melambat secara signifikan dan membentuk sudut siku (*elbow*) menandakan banyak klaster yang optimal [10].

2.5 Analisis Klaster

Analisis klaster merupakan pengelompokan objek berdasarkan karakteristik tertentu, di mana objek-objek tersebut dikelompokkan ke dalam kategori yang telah ditentukan [11]. Dengan demikian, analisis klaster mengelompokkan objek sedemikian sehingga objek yang mewakili kesamaan paling tinggi dengan objek lainnya akan berada dalam klaster yang sama [12]. Menurut [13] klaster yang baik memiliki tingkat homogenitas yang tinggi di antara anggota dalam satu klaster (*within-cluster*) dan memiliki heterogenitas yang tinggi di antara klaster yang berbeda (*between cluster*). Terdapat dua metode dalam analisis klaster, yaitu:

1. Metode Hierarki (*Hierarchical Method*)

Metode hierarki merupakan metode pengelompokan dengan menggabungkan dua atau lebih objek yang memiliki kemiripan tertinggi dan menampilkan hasilnya dalam bentuk dendrogram [13].

2. Metode Non-Hierarki (*Non-Hierarchical Method*)

Metode Non-Hierarki adalah metode pengelompokan yang menentukan banyak kluster di awal sebelum memperbarui pusat kluster berdasarkan jarak antar objek hingga mencapai kriteria tertentu [6], [14].

2.6 K-Means Clustering

K-means clustering adalah metode pengelompokan non-hierarki yang digunakan untuk mengelompokkan data ke dalam dua kelompok atau lebih. Tujuan dari metode ini adalah untuk meminimalkan kesamaan di dalam setiap kelompok dan memaksimalkan perbedaan di antara kelompok-kelompok tersebut. Metode ini membagi data menjadi sejumlah kluster yang telah ditentukan di awal berdasarkan jarak antar data. *K-means clustering* banyak diterapkan karena kemampuannya dalam mengelompokkan data besar dan *outlier* dengan cepat dan efisien, terutama pada data yang bersifat numerik [5].

Dalam *K-means clustering*, setiap kluster diwakili oleh titik pusat atau *centroid*, yang diperbarui secara berkala hingga mencapai hasil pengelompokan yang optimal. Prosesnya dilakukan melalui iterasi berulang, di mana data dikelompokkan ke dalam kategori yang paling sesuai [15]. Metode ini menggunakan jarak *Euclidean* untuk mengukur ketidakmiripan antara data dan *centroid*, sehingga dapat menentukan keanggotaan objek dalam kluster tertentu [16].

3. Hasil dan Pembahasan

3.1 Statistik Deskriptif

Hasil perhitungan statistik deskriptif dari data pengangguran terbuka berdasarkan pendidikan tertinggi yang ditamatkan di Provinsi Jawa Barat dengan menggunakan *software IBM SPSS Statistic 26* tersaji pada Tabel 1.

Tabel 1. Statistik Deskriptif

	N	Minimum	Maximum	Mean	Std. Deviation
SD ke bawah	27	662	52.659	13.466,74	11.568,152
SMP	27	513	39.042	11.645,15	9.096,238
SMA	27	703	55.788	19.981,22	15.469,787
SMK	27	1654	67.286	19.481,19	14.324,936
Diploma/Akademi/Universitas	27	0	23.868	5.362,26	6.184,876
Valid N (listwise)	27				

Berdasarkan *output* tersebut dapat diketahui statistik deskriptif dari variabel SD ke bawah, SMP, SMA, SMK, dan diploma/akademi/universitas. Statistik deskriptif tersebut memuat banyak data *N*, nilai minimum, nilai maksimum, rata-rata, dan standar deviasi. Terdapat 27 data kabupaten/kota di Provinsi Jawa Barat yang akan digunakan pada penelitian ini.

3.2 Standarisasi Data dan Deteksi Outliner

Standarisasi data dilakukan untuk memastikan bahwa rentang nilai data tidak terlalu bervariasi, sehingga pengelompokan data dapat dilakukan dengan baik. Proses standarisasi data dilakukan dengan mengubah data menjadi nilai standar atau *z-score*. Data dengan nilai *z-score* lebih dari 3,00 atau kurang dari -3,00 dianggap sebagai data

yang menyimpang (*outlier*). Hasil perhitungan nilai *z-score* dengan menggunakan *software* IBM SPSS *Statistic* 26 disajikan pada Tabel 2.

Tabel 2. Nilai *Z-Score* dari Data Pengangguran Terbuka

No.	Kabupaten/Kota	<i>Z-Score</i>				
		SD ke bawah	SMP	SMA	SMK	Diploma/Akademi/Universitas
1	Kabupaten Bogor	3,38794	3,01189	2,31463	3,33717	1,86758
2	Kabupaten Sukabumi	0,35730	1,35582	0,90989	0,36152	0,09244
3	Kabupaten Cianjur	1,17022	1,32493	0,77104	0,21639	-0,53975
4	Kabupaten Bandung	0,96111	1,14221	1,49296	0,38079	0,39350
5	Kabupaten Garut	1,23756	0,36343	0,47278	0,51064	-0,02866
6	Kabupaten Tasikmalaya	-0,61667	-0,38688	-0,84909	-0,15638	-0,75333
7	Kabupaten Ciamis	-0,94291	-0,87082	-0,52484	-1,01119	-0,86700
8	Kabupaten Kuningan	-0,09187	-0,09929	-0,80151	0,11001	-0,42624
9	Kabupaten Cirebon	0,36032	0,63959	0,44673	0,42924	-0,27749
10	Kabupaten Majalengka	-0,60466	-0,71899	-1,04696	-0,67534	-0,13909
11	Kabupaten Sumedang	-0,49193	-0,13788	-0,26770	-0,76532	-0,71178
12	Kabupaten Indramayu	-0,46181	0,00361	-0,29937	-0,43583	-0,57596
13	Kabupaten Subang	-0,09360	0,42115	-0,33260	0,41276	-0,62544
14	Kabupaten Purwakarta	-0,42304	-0,50550	-0,78496	-0,21977	-0,81720
15	Kabupaten Karawang	-0,13293	1,13111	0,86781	1,34240	-0,23416
16	Kabupaten Bekasi	1,51219	0,95356	0,99606	1,79392	0,65430
17	Kabupaten Bandung Barat	-0,13250	-0,28365	0,88351	-0,14103	-0,09964
18	Kabupaten Pangandaran	-1,10413	-1,22382	-1,24619	-1,23513	-0,79165
19	Kota Bogor	-0,70234	-0,52232	-0,41612	-0,04741	0,04103
20	Kota Sukabumi	-1,01846	-1,01483	-1,03778	-1,00693	-0,61865
21	Kota Bandung	0,50736	-0,06829	1,66562	0,07461	2,32984
22	Kota Cirebon	-0,92718	-1,13279	-1,09499	-1,06466	-0,46101
23	Kota Bekasi	-0,31844	-0,03695	0,91868	0,38645	2,99210
24	Kota Depok	-0,14088	-0,32103	-0,20428	0,15587	1,22618
25	Kota Cimahi	-0,50395	-0,78551	-0,74702	-0,70152	-0,34993
26	Kota Tasikmalaya	-0,60344	-1,02615	-0,93636	-0,80679	-0,54088
27	Kota Banjar	-1,10690	-1,21261	-1,14993	-1,24449	-0,73910

Berdasarkan nilai *z-score* yang terdapat pada Tabel 2, terdeteksi satu *outlier*, yaitu data dari Kabupaten Bogor yang memiliki nilai *z-score* lebih dari 3,00. Jika *outlier* tidak dibuang, maka pembagian kluster menjadi tidak seimbang dimana kluster 1 hanya memuat *outlier* saja. Oleh karena itu, perlu dilakukan penghapusan terhadap *outlier* tersebut. Dengan demikian, yang semula banyak datanya ada 27 menjadi 26 data.

3.3 Uji Statistik Hopkins

Uji statistik Hopkins adalah metode yang digunakan untuk memeriksa validasi data memiliki kecenderungan membentuk kluster. Berdasarkan hasil uji statistik Hopkins menggunakan *software* *RStudio* diperoleh nilai 0,9366731. Karena nilainya lebih dari 0,5, maka nilai tersebut menunjukkan kecenderungan data untuk membentuk kluster. Oleh karena itu, pengelompokan dapat dilakukan pada data pengangguran terbuka.

3.4 Uji Asumsi Klasik Analisis Kluster

Dalam analisis kluster, terdapat dua asumsi yang perlu dipenuhi sebagai berikut:

1. Sampel data yang diambil harus dapat mewakili populasi dari suatu data. Untuk menentukan apakah sampel tersebut sudah mewakili populasi dan dapat digunakan untuk analisis lebih lanjut, diperlukan pengujian menggunakan uji *Kaiser Meyer Olkin* (KMO), yang digunakan untuk mengukur korelasi antar variabel yang dapat memengaruhi hasil *clustering*. Nilai KMO yang diperoleh melalui perhitungan menggunakan *software IBM SPSS Statistics 26* tersaji pada Tabel 3.

Tabel 3. Nilai KMO

<i>Kaiser-Meyer-Olkin Measure of Sampling Adequacy.</i>		0,688
<i>Bartlett's Test of Sphericity</i>	<i>Approx. Chi-Square</i>	102,989
	<i>df</i>	10
	<i>Sig.</i>	0,000

Berdasarkan Tabel 3, nilai KMO yang diperoleh adalah 0,688. Karena nilai KMO lebih dari 0,5, maka data tersebut dianggap sudah cukup mewakili untuk populasi yang diteliti dan dapat digunakan dalam analisis lebih lanjut. Dengan demikian, akan menghasilkan *clustering* yang lebih stabil dan mudah diinterpretasi.

2. Asumsi non-multikolinieritas diperlukan dalam *K-means* agar interpretasi kluster menjadi lebih jelas dan menghasilkan *clustering* yang konsisten. Hal ini akan mengakibatkan hasil *clustering* yang lebih akurat. Multikolinieritas dapat dideteksi melalui nilai *tolerance* dan *Variance Inflation Factor* (VIF). Tabel 4 hingga Tabel 8 menyajikan hasil analisis multikolinieritas menggunakan *software IBM SPSS Statistics 26*.

Tabel 4. Korelasi Variabel SD ke Bawah dengan Variabel Lainnya

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
	(Constant)	673,279	1.885,733				0,357
SMP	0,485	0,364	0,414	1,331	0,197	0,134	7,483
SMA	0,207	0,175	0,334	1,187	0,249	0,163	6,141
SMK	0,158	0,186	0,199	0,854	0,403	0,239	4,193
Diploma/Akademi/Universitas	-0,103	0,281	-0,069	-0,365	0,719	0,359	2,786

Tabel 5. Korelasi Variabel SMP dengan Variabel Lainnya

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
	(Constant)	198,667	1.087,071				0,183
SD ke Bawah	0,160	0,121	0,188	1,331	0,197	0,294	3,400
SMA	0,288	0,083	0,544	3,482	0,002	0,241	4,155
SMK	0,300	0,087	0,441	3,467	0,002	0,362	2,759
Diploma/Akademi/Universitas	-0,443	0,130	-0,350	-3,399	0,003	0,553	1,809

Tabel 6. Korelasi Variabel SMA dengan Variabel Lainnya

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
	(Constant)	-180,836	2.285,561				-0,079
SD ke Bawah	0,303	0,255	0,188	1,187	0,249	0,289	3,456
SMP	1,271	0,365	0,673	3,482	0,002	0,194	5,145
SMK	-0,221	0,223	-0,172	-0,989	0,334	0,241	4,145

Diploma/Akademi/Universitas	1,137	0,234	0,476	4,855	0,000	0,757	1,321
-----------------------------	-------	-------	-------	-------	-------	-------	-------

Tabel 7. Korelasi Variabel SMK dengan Variabel Lainnya

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	2.946,276	2.089.684		1,410	0,173		
SD ke Bawah	0,212	0,248	0,169	0,854	0,403	0,281	3,564
SMP	1,213	0,350	0,825	3,467	0,002	0,194	5,160
SMA	-0,202	0,204	-0,259	-0,989	0,334	0,160	6,261
Diploma/Akademi/Universitas	0,626	0,296	0,336	2,110	0,047	0,432	2,313

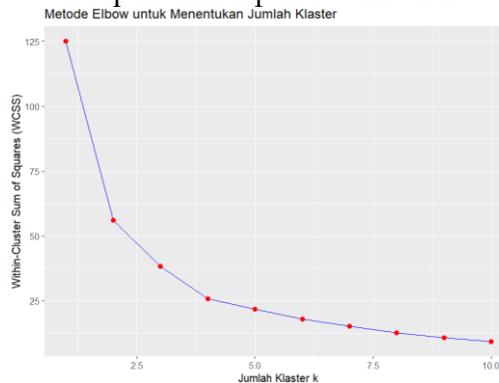
Tabel 8. Korelasi Variabel Diploma/Akademi/Universitas dengan Variabel Lainnya

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	542,388	1.457,336		0,372	0,713		
SD ke Bawah	-0,061	0,168	-0,091	-0,365	0,719	0,273	3,664
SMP	-0,801	0,236	-1,013	-3,399	0,003	0,191	5,235
SMA	0,465	0,096	1,112	4,855	0,000	0,324	3,087
Diploma/Akademi/Universitas	0,280	0,133	0,520	2,110	0,047	0,279	3,579

Berdasarkan hasil perhitungan dengan *software* IBM SPSS Statistics 26 diperoleh bahwa semua variabel memiliki nilai VIF kurang dari 10 dan nilai *tolerance* lebih besar dari 0,1. Oleh karena itu, semua variabel tidak terjadi multikolinieritas. Dengan demikian, dapat dilakukan analisis *K-means clustering*.

3.5 Menentukan Banyak Kluster

Terdapat beberapa metode untuk menentukan banyak kluster optimal dalam *K-means*. Pada penelitian ini, digunakan metode *elbow* untuk tujuan tersebut. Prosesnya dilakukan dengan menghitung nilai *Within-cluster Sum of Squares* (WCSS), lalu hasilnya divisualisasikan dalam bentuk kurva. Kurva WCSS membantu menentukan banyak kluster optimal yaitu dengan melihat titik yang membentuk siku (*elbow*) saat penurunan nilai WCSS mulai melambat secara signifikan. Visualisasi hasil metode *elbow* yang dibuat menggunakan *software RStudio* dapat dilihat pada Gambar 1.



Gambar 1. Hasil Pengujian Metode *Elbow*

Berdasarkan Gambar 1, diperoleh bahwa banyak kluster optimal adalah 3. Hal ini ditunjukkan oleh titik *elbow* pada kurva di atas yang menandakan penurunan signifikan dalam WCSS hingga $k = 3$, setelah itu penurunan menjadi lebih lambat. Dengan demikian, pada penelitian ini akan dibentuk 3 kluster yang meliputi kluster 1, kluster 2, dan kluster 3.

3.6 Penentuan Pusat Kluster (*Centroid*) Awal

Penentuan pusat kluster (*centroid*) awal dapat dilakukan secara acak dari data pengangguran terbuka. Langkah tersebut memungkinkan setiap titik data memiliki kesempatan yang sama untuk menjadi *centroid* awal. Penentuan *centroid* awal dilakukan menggunakan *software* IBM SPSS Statistics 26. Hasil dari pemilihan acak tersebut disajikan dalam Tabel 9, yang menampilkan nilai-nilai *centroid* awal untuk setiap kluster yang terbentuk.

Tabel 9. Nilai *Centroid* Awal

<i>Z-score</i>	<i>Cluster</i>		
	1	2	3
SD ke Bawah	-1,10413	1,51219	-0,31844
SMP	-1,22382	0,95356	-0,03695
SMA	-1,24619	0,99606	0,91868
SMK	-1,23513	1,79392	0,38645
Diploma/Akademi/Universitas	-0,79165	0,65430	2,99210

Berdasarkan Tabel 9, dapat dilihat bahwa nilai *centroid* awal pada kluster 1 merupakan nilai *z-score* Kabupaten Pangandaran, kluster 2 merupakan nilai *z-score* Kabupaten Bekasi, dan kluster 3 merupakan nilai *z-score* Kota Bekasi.

3.7 Proses Iterasi

Iterasi optimal dapat diidentifikasi secara lebih efisien dengan menggunakan *software* IBM SPSS Statistics 26. Jumlah iterasi optimal dapat dilihat pada Tabel 10.

Tabel 10. Jumlah Iterasi Optimal

<i>Iteration</i>	<i>Change in Cluster Centers</i>		
	1	2	3
1	1,247	1,459	0,909
2	0,000	0,000	0,000

- a. *Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is 0,000. The current iteration is 2. The minimum distance between initial centers is 3,433.*

Berdasarkan Tabel 10 menunjukkan proses iterasi dalam analisis kluster. Pada iterasi pertama terjadi perubahan signifikan pada *centroid*, di mana kluster 1, 2, dan 3 masing-masing mengalami perubahan sebesar 1,247, 1,459, dan 0,909. Namun, pada iterasi kedua nilai perubahan *centroid* untuk ketiga kluster tersebut menjadi 0,000 yang menandakan bahwa tidak ada lagi perubahan posisi *centroid* dan proses pengelompokan (*clustering*) mencapai kestabilan.

Kestabilan tercapai karena tidak ada lagi perubahan pada *centroid*. Perubahan terbesar pada *centroid* adalah 0,000 sehingga proses iterasi berhenti pada iterasi kedua.

Jarak minimum antar *centroid* berdasarkan hasil iterasi yaitu 3,433. Dengan tercapainya kestabilan pada iterasi kedua, perhitungan ini dianggap akurat dan dapat dilanjutkan ke tahap interpretasi kluster serta analisis karakteristiknya.

3.8. Interpretasi dan Karakteristik Kluster

Karakteristik setiap kluster dapat dilihat dari rata-rata nilai (*centroid*) dari setiap variabel untuk masing-masing kluster setelah proses iterasi selesai. Setelah iterasi selesai, *centroid* tersebut digunakan untuk memahami pola dan hubungan antar kabupaten/kota dalam setiap kluster, serta membandingkan karakteristik antar kluster secara lebih rinci. Hal tersebut disajikan dalam Tabel 11 yang diperoleh melalui penggunaan *software* IBM SPSS *Statistic* 26.

Tabel 11. Final Cluster Centers

Z-score	Cluster		
	1	2	3
SD ke Bawah	-0,55636	0,78082	0,01601
SMP	-0,59352	0,98724	-0,14209
SMA	-0,66574	0,85104	0,79334
SMK	-0,56181	0,71927	0,20564
Diploma/Akademi/Universitas	-0,52974	0,00860	2,18270

Berdasarkan *output* pada Tabel 11, data tersebut masih terkait dengan proses standarisasi menggunakan *z-score*. Kolom *Cluster* berisi nilai rata-rata pengangguran terbuka berdasarkan pendidikan tertinggi yang ditamatkan. Nilai positif menunjukkan bahwa rata-rata variabel dalam kluster lebih tinggi dari rata-rata keseluruhan populasi, sedangkan nilai negatif menunjukkan bahwa rata-rata variabel lebih rendah dari rata-rata keseluruhan populasi. Berikut adalah interpretasi karakteristik dari setiap kluster.

1. Kluster 1

Semua nilai pada kluster 1 bernilai negatif yang menunjukkan bahwa rata-rata pengangguran terbuka di kluster ini berada di bawah rata-rata total. Oleh karena itu, kluster 1 mewakili kelompok kabupaten/kota dengan tingkat pengangguran terbuka rendah berdasarkan pendidikan tertinggi yang ditamatkan.

2. Kluster 2

Semua nilai pada kluster 2 bernilai positif yang menunjukkan bahwa rata-rata pengangguran terbuka di kluster ini berada di atas rata-rata total. Oleh karena itu, kluster 2 mewakili kelompok kabupaten/kota dengan pengangguran terbuka tinggi berdasarkan pendidikan tertinggi yang ditamatkan.

3. Kluster 3

Nilai rata-rata pengangguran terbuka SMP berada di bawah rata-rata total, sedangkan untuk SD ke bawah, SMA, SMK, serta Diploma/Akademi/Universitas berada di atas rata-rata total. Oleh karena itu, kluster 3 mewakili kelompok kabupaten/kota dengan pengangguran terbuka sedang berdasarkan pendidikan tertinggi yang ditamatkan.

Sebagai tambahan, subjek dari setiap kluster di atas merupakan penduduk usia 15 tahun ke atas. Setelah tiga kluster terbentuk, langkah selanjutnya adalah menganalisis perbedaan pada variabel-variabel dalam setiap kluster. Perbedaan ini dapat dianalisis menggunakan nilai F dan nilai probabilitas (*sig*) untuk setiap variabel. Hal ini dilakukan dengan mengacu pada *output* ANOVA (*Analysis of Variance*) yang terdapat di Tabel 12.

Tabel 12. ANOVA

	Cluster		Error		F	Sig.
	Mean Square	Df	Mean Square	Df		
Zscore (SD ke Bawah)	4,390	2	0,230	23	19,048	0,000
Zscore (SMP)	6,085	2	0,192	23	31,742	0,000
Zscore (SMA)	6,922	2	0,287	23	24,146	0,000
Zscore (SMK)	4,185	2	0,264	23	15,871	0,000
Zscore (Diploma/Akademi/Universitas)	9,324	2	0,162	23	57,511	0,000

Berdasarkan Tabel 12, terdapat perbedaan signifikan antara kluster 1, 2, dan 3, karena semua variabel memiliki nilai F yang besar dan nilai sig kurang dari 0,05. Di antara variabel-variabel tersebut, variabel diploma/akademi/universitas menunjukkan perbedaan terbesar dengan nilai F tertinggi sebesar 57,511 dan sig 0,000 yang mengindikasikan signifikansi nyata. Oleh karena itu, perbedaan karakteristik antar kluster 1, 2, dan 3 disebabkan oleh variasi pada semua variabel.

Setelah menganalisis variabel yang memberikan perbedaan pada kluster yang terbentuk, akan didapatkan banyak anggota dari setiap kluster. Berikut disajikan banyak anggota dari setiap kluster berdasarkan hasil *clustering* menggunakan *software* IBM SPSS *Statistic* 26.

Tabel 13. Banyak Anggota Setiap Kluster

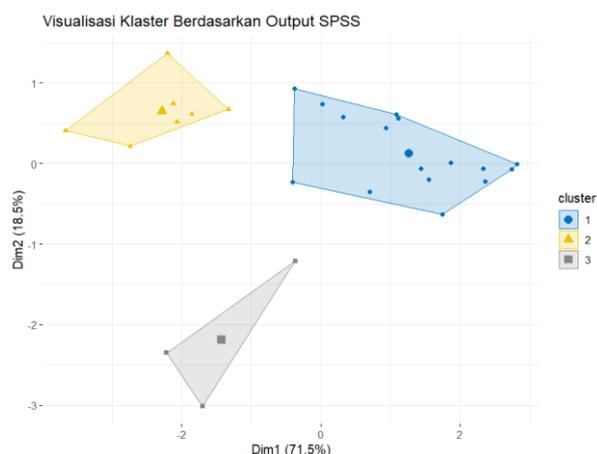
Cluster	Cluster			Valid	Missing
	1	2	3		
	16,000	7,000	3,000	26,000	0,000

Berikut adalah detail mengenai anggota dari kluster 1, kluster 2, dan kluster 3 yang telah dibentuk.

1. Kluster 1 memiliki 16 anggota, meliputi 10 kabupaten (Tasikmalaya, Ciamis, Kuningan, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Bandung Barat, dan Pangandaran) dan 6 kota (Bogor, Sukabumi, Cirebon, Cimahi, Tasikmalaya, dan Banjar).
2. Kluster 2 memiliki 7 kabupaten sebagai anggotanya (Cirebon, Karawang, Bekasi, Sukabumi, Cianjur, Bandung, dan Garut).
3. Kluster 3 memiliki 3 kota sebagai anggotanya (Bandung, Bekasi, dan Depok).

Pembagian menjadi tiga kluster di atas menunjukkan bahwa permasalahan pengangguran di setiap wilayah memiliki karakteristik yang berbeda. Diperoleh bahwa kluster 1 memiliki tingkat pengangguran terbuka yang rendah, terutama pada kelompok penduduk dengan pendidikan tinggi. Ini menunjukkan bahwa di wilayah kluster 1, tingkat pendidikan yang lebih tinggi berkorelasi dengan peluang kerja yang lebih baik. Di sisi lain, kluster 2 yang memiliki tingkat pengangguran terbuka yang tinggi, terutama pada kelompok penduduk dengan pendidikan rendah. Hal ini menunjukkan adanya kesenjangan yang lebih besar antara persediaan tenaga kerja dengan permintaan pasar kerja di wilayah kluster 2. Sementara itu, pada kluster 3 dengan tingkat pengangguran

terbuka yang sedang. Klaster 3 menunjukkan kondisi yang lebih beragam, dengan beberapa kabupaten/kota memiliki tingkat pengangguran yang lebih tinggi pada kelompok pendidikan tertentu, sementara yang lainnya lebih rendah. Lebih lanjut, anggota dari klaster 3 merupakan pusat pertumbuhan ekonomi dan pendidikan yang menarik imigran lulusan dari berbagai daerah. Namun, peningkatan jumlah tenaga kerja terdidik tidak berbanding lurus dengan ketersediaan lapangan kerja yang sesuai. Urbanisasi yang pesat tanpa perencanaan ketenagakerjaan yang memadai turut memperbesar kesenjangan antara jumlah pencari kerja dan permintaan tenaga kerja, khususnya di sektor formal yang menyerap tenaga kerja secara selektif. Untuk memperjelas hasil pengelompokan, visualisasi klasterisasi kabupaten/kota berdasarkan pengangguran terbuka menurut jenjang pendidikan yang ditamatkan telah tersaji pada Gambar 2.



Gambar 2. Visualisasi Klaster Pengangguran Terbuka

3.9. Validasi Klaster Akhir

Validasi klaster dilakukan dengan menghitung nilai *Silhouette Coefficient* untuk menilai kualitas hasil pengelompokan. Hasil perhitungan di RStudio menunjukkan rata-rata nilai sebesar 0,4751387. Nilai ini termasuk dalam kategori *medium structure* yang berarti sebagian besar data sudah tergolong dalam klaster yang tepat. Namun, masih ada kemungkinan beberapa data kurang sesuai sehingga metode tambahan dapat dipertimbangkan untuk memperjelas keanggotaan klaster.

4. Kesimpulan

Berdasarkan pembahasan yang telah diberikan, pengelompokan pengangguran terbuka Provinsi Jawa Barat menghasilkan tiga klaster, di mana klaster 1 (16 anggota), klaster 2 (7 anggota), dan klaster 3 (3 anggota) secara berturut-turut merupakan klaster dengan tingkat pengangguran terbuka yang rendah, tinggi, dan sedang. Pada pembagian tersebut, wilayah dengan tingkat pendidikan penduduk yang lebih tinggi cenderung memiliki tingkat pengangguran yang relatif lebih rendah. Dengan demikian, diperlukan kebijakan yang terintegrasi dan disesuaikan dengan kondisi spesifik masing-masing wilayah, yang diharapkan dapat mengurangi tingkat pengangguran dan meningkatkan kesejahteraan masyarakat Jawa Barat.

Penelitian lebih lanjut dapat dilakukan dengan menyertakan faktor lain yang memengaruhi pengangguran terbuka, seperti kondisi ekonomi, usia, dan jenis kelamin. Selain itu, penerapan metode *clustering* maupun metode penentuan banyak kluster optimal lain dengan data yang sama juga dapat dilakukan. Dengan membandingkan hasil *clustering* dari metode lain dengan metode *K-means clustering*, diharapkan dapat diperoleh hasil yang lebih baik dan akurat.

Daftar Pustaka

- [1] Sarbaini, W. Saputri, Nazaruddin, and F. Muttakin, "Cluster Analysis Menggunakan Algoritma Fuzzy K-Means untuk Tingkat Pengangguran di Provinsi Riau," *Jurnal Teknologi dan Manajemen Industri Terapan (JTMIT)*, vol. 1, no. II, pp. 78–84, 2022.
- [2] A. Aziz and S. Norhasanah, "Klastrisasi Data Tenaga Kerja Terbuka Menurut Provinsi dengan Penggunaan Algoritma K-Means," 2023. [Online]. Available: <http://jurnal.mdp.ac.id>
- [3] T. Firmansyah and Yuliazmi, "Implementasi Klasterisasi K-Means untuk Memetakan Jumlah Pengangguran di Provinsi Jawa Barat Berdasarkan Tingkat Pendidikan," *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi*, vol. 2, no. 2, pp. 995–1003, 2023.
- [4] R. A. Nugraheni and R. Mukti, *Keadaan Angkatan Kerja di Provinsi Jawa Barat*. Badan Pusat Statistik Provinsi Jawa Barat, 2024.
- [5] A. Mahmudan, "Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering (Pengelompokan Kabupaten/Kota di Jawa Tengah Berdasarkan Kasus COVID-19 Menggunakan K-Means Clustering)," *Jurnal Matematika, Statistika, & Komputasi*, vol. 17, no. 1, pp. 1–13, 2020, doi: 10.20956/jmsk.v%vi%i.10727.
- [6] R. N. Puspita, "Analisis K-Means Cluster pada Kabupaten/Kota di Provinsi Banten Berdasarkan Indikator Indeks Pembangunan Manusia," *Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 2, no. 3, pp. 267–281, 2021, doi: 10.46306/lb.v2i3.
- [7] S. K. Arum, R. Astuti, and F. M. Basysyar, "Penerapan Algoritma K-Means pada Dataset Pengangguran Terbuka Berdasarkan Pendidikan di Provinsi Jawa Barat," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 2, pp. 2221–2226, 2024.
- [8] R. H. Syah, "Penerapan Metode K-Means dalam Pengelompokan Kabupaten/Kota di Jawa Tengah (Berdasarkan Tingkat Kerentanan Tanah Longsor Tahun 2017)," *Jurnal Ilmiah Populer*, vol. 01, pp. 44–58, 2019.
- [9] N. P. E. U. Barsua, I. M. J. J. Dilaga, R. L. Simbolon, and R. Kurniawan, "Pengelompokan Toko Pupuk Termurah E-commerce Shopee dengan Metode Klasterisasi," *Seminar Nasional Sains Data (SENADA)*, 2024.
- [10] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *Journal of Artificial Intelligence and Technology Information (JAITI)*, vol. 1, no. 1, pp. 20–28, 2023, doi: 10.58602/jaiti.v1i1.22.
- [11] F. A. Ulya, A. N. Abdullah, T. A. Hanan, and I. M. Nur, "Pengelompokan Tingkat Pengangguran Terbuka di Jawa Tengah Menggunakan Metode K-Means Clustering," *Journal of Data Insights*, vol. 1, no. 2, pp. 71–80, 2023, doi: 10.26714/jodi.
- [12] M. Rais, R. Goejantoro, and S. Prangga, "Optimalisasi K-Means Cluster dengan Principal Component Analysis pada Pengelompokan Kabupaten/Kota di Pulau Kalimantan Berdasarkan Indikator Tingkat Pengangguran Terbuka," *Jurnal EKSPONENSIAL*, vol. 12, no. 2, 2021.
- [13] S. Santoso, *Statistik Multivariat dengan SPSS*. Jakarta: PT Elex Media Komputindo, 2017.
- [14] Wardono, Sunarmi, and M. R. Wirawan, "Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan dengan Metode K-Means Cluster," *Seminar Nasional Edusainstek*, pp. 599–610, 2019, [Online]. Available: <http://prosidings.unimus.ac.id>
- [15] N. P. E. Merliana, Ernawati, and A. J. Santoso, *Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering*. 2015.
- [16] Athifaturrofifah, R. Goejantoro, and D. Yuniarti, "Perbandingan Pengelompokan K-Means dan K-Medoids pada Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Studi Kasus: Data Titik Panas di Indonesia pada 28 April 2018)," *Jurnal Eksponensial*, vol. 10, no. 2, 2019.