

# KLASTERISASI KABUPATEN/KOTA DI JAWA BARAT BERDASARKAN KUALIFIKASI DAN PEMERATAAN GURU SMP MENGGUNAKAN K-MEANS CLUSTERING

Fitri Rahmawati<sup>1\*</sup>, Nurmala Setianing Putri<sup>2</sup>, Nadia Ulfa<sup>2</sup>, Ulfah Nur Azizah<sup>2</sup>

<sup>1</sup>Program Studi Matematika, FPMIPA, Universitas Pendidikan Indonesia

<sup>2</sup>Program Studi Pendidikan Matematika, FPMIPA, Universitas Pendidikan Indonesia

\*fitrirahmawati@upi.edu

## ABSTRAK

Penelitian ini bertujuan untuk mengklasterisasi kabupaten di Jawa Barat berdasarkan kualifikasi dan pemerataan guru SMP menggunakan metode K-Means Clustering. Variabel yang dianalisis mencakup kualifikasi guru, guru tersertifikasi, dan Indeks Pemerataan Guru (IPG) PNS. Kualifikasi guru menunjukkan persentase guru yang telah memenuhi standar pendidikan S1/D4, sertifikasi guru menggambarkan persentase guru yang telah tersertifikasi sebagai bukti pemenuhan standar profesionalisme, sementara IPG PNS digunakan untuk mengukur tingkat pemerataan distribusi guru PNS di wilayah Kabupaten/Kota. Hasil analisis K-Means Clustering membagi kabupaten/kota di Jawa Barat menjadi dua klaster. Klaster 1, yang terdiri dari 12 Kabupaten/Kota, menunjukkan kualitas guru yang cukup baik dengan persentase kualifikasi guru dan guru tersertifikasi di atas rata-rata, serta distribusi guru yang lebih merata. Sebaliknya, Klaster 2, yang mencakup 15 Kabupaten/Kota, memiliki persentase kualifikasi guru dan guru tersertifikasi di bawah rata-rata, mengindikasikan kualitas guru yang masih kurang serta distribusi guru yang kurang merata. Temuan ini memberikan gambaran mengenai perbedaan kondisi tenaga pendidik di Jawa Barat, yang dapat dijadikan dasar dalam penyusunan kebijakan peningkatan kualitas dan pemerataan guru.

**Kata kunci:** K-Means Clustering; Kualifikasi Guru; Sertifikasi Guru; Indeks Pemerataan Guru

## ABSTRACT

*This study aims to cluster regencies in West Java based on junior high school teacher qualifications and distribution using the K-Means Clustering method. The analyzed variables include teacher qualifications, certified teachers, and the Civil Servant Teacher Distribution Index (IPG PNS). Teacher qualifications refer to the percentage of teachers who meet the educational standard of a bachelor's degree (S1/D4), certified teachers indicate the percentage of teachers certified as proof of meeting professional standards, while IPG PNS measures the level of equity in the distribution of civil servant teachers across regencies/cities. The K-Means Clustering analysis divided the regencies/cities in West Java into two clusters. Cluster 1, consisting of 12 regencies/cities, demonstrates relatively good teacher quality, with higher-than-average percentages of qualified and certified teachers, as well as a more equitable teacher distribution. In contrast, Cluster 2, comprising 15 regencies/cities, shows lower-than-average percentages of qualified and certified teachers, indicating lower teacher quality and less equitable teacher distribution. These findings provide insights into disparities in teacher conditions in West Java, which can serve as a basis for formulating policies to improve teacher quality and distribution.*

**Key words:** K-Means Clustering; Teacher Qualifications; Teacher Certification; Teacher Distribution Index

## PENDAHULUAN

Pendidikan merupakan salah satu elemen penting dalam pembangunan sumber daya manusia yang berkualitas. Dalam konteks pendidikan di Indonesia, keberadaan guru yang kompeten dan distribusi yang merata menjadi kunci utama untuk memastikan setiap peserta didik mendapatkan hak yang sama atas pendidikan berkualitas. Pemerataan guru, khususnya di tingkat Sekolah Menengah Pertama (SMP), sering kali menjadi tantangan besar bagi pemerintah, terutama di wilayah-wilayah dengan karakteristik geografis dan sosial yang beragam seperti Provinsi Jawa Barat. Tantangan ini

semakin relevan mengingat Jawa Barat merupakan provinsi dengan jumlah penduduk terbesar di Indonesia, yang secara otomatis memiliki kebutuhan pendidikan yang kompleks.

Ketimpangan distribusi guru sering kali menjadi salah satu penyebab utama perbedaan kualitas pendidikan antarwilayah. Kabupaten-kabupaten tertentu cenderung memiliki kelebihan guru dengan kualifikasi memadai, sementara kabupaten lainnya justru mengalami kekurangan guru berkualitas. Ketidakeimbangan ini dapat berdampak negatif terhadap capaian pembelajaran siswa, menciptakan kesenjangan pendidikan yang semakin sulit diatasi. Dalam menghadapi masalah ini, penting untuk memahami pola distribusi guru dan faktor-faktor yang memengaruhinya agar dapat dirumuskan kebijakan yang tepat untuk pemerataan dan peningkatan kualitas pendidikan di seluruh wilayah Jawa Barat.

Penelitian sebelumnya telah menunjukkan pentingnya analisis berbasis data dalam mengkaji pemerataan pendidikan. (Suputra et al., 2021) menggunakan metode klasterisasi untuk mengelompokkan provinsi di Indonesia berdasarkan nilai ujian nasional, yang memberikan wawasan tentang pemerataan pendidikan di berbagai wilayah. (Putri & Abdullah, 2021) menerapkan analisis klastering untuk mengelompokkan Sekolah Menengah Atas (SMA) di Provinsi Jawa Barat berdasarkan ketersediaan fasilitas. Lebih lanjut, (Wibowo & Habanabakize, 2022) melakukan analisis klastering terhadap standar kualifikasi pendidikan dan pengalaman kerja guru Sekolah Menengah Kejuruan (SMK) di Indonesia. Meskipun telah banyak analisis klastering mengenai pemerataan pendidikan yang telah dilakukan sebelumnya, penelitian yang secara spesifik mengkaji distribusi dan kualifikasi guru SMP di kabupaten/kota di Jawa Barat masih terbatas, sehingga diperlukan analisis lebih lanjut untuk mengisi kesenjangan ini.

Dalam konteks analisis data, metode klastering menjadi salah satu pendekatan yang efektif untuk mengelompokkan data berdasarkan karakteristik tertentu. Clustering adalah teknik dalam data mining yang bertujuan untuk mengidentifikasi kelompok-kelompok data yang memiliki kesamaan tertentu. Salah satu algoritma yang paling populer digunakan dalam klastering adalah K-Means Clustering. Metode ini bekerja dengan membagi data ke dalam sejumlah klaster berdasarkan jarak antar data, sehingga setiap klaster merepresentasikan kelompok dengan karakteristik yang serupa. Dalam konteks penelitian ini, K-Means Clustering digunakan untuk mengelompokkan kabupaten di Jawa Barat berdasarkan kualifikasi dan pemerataan guru SMP. Pendekatan ini memberikan gambaran visual dan kuantitatif mengenai pola distribusi sumber daya pendidikan, yang sangat berguna dalam proses pengambilan keputusan.

Penelitian ini menawarkan kebaruan dalam pendekatannya dengan menggunakan metode K-Means Clustering untuk menganalisis pemerataan dan kualifikasi guru SMP di Jawa Barat. Metode ini memungkinkan pengelompokan kabupaten berdasarkan kesamaan karakteristik, seperti jumlah guru yang berkualifikasi, rasio guru terhadap siswa, dan indikator pemerataan lainnya. Hasil dari analisis ini diharapkan dapat memberikan gambaran yang lebih jelas mengenai distribusi sumber daya pendidikan di Jawa Barat, serta menjadi dasar dalam perumusan kebijakan yang lebih efektif dan berkelanjutan. Selain memberikan kontribusi teoretis, penelitian ini juga memiliki nilai praktis yang signifikan. Dengan mengidentifikasi pola-pola distribusi dan kualifikasi guru, pemerintah daerah dapat mengalokasikan sumber daya pendidikan secara lebih strategis, mengurangi kesenjangan antarwilayah, dan meningkatkan kualitas pendidikan secara keseluruhan.

Hasil penelitian ini juga dapat dimanfaatkan oleh pemangku kebijakan di tingkat nasional sebagai contoh aplikasi analisis data untuk perencanaan kebijakan pendidikan di wilayah lain di Indonesia. Dengan pendekatan ini, penelitian diharapkan dapat menjawab beberapa pertanyaan penting, seperti:

bagaimana pola distribusi guru SMP di kabupaten/kota di Jawa Barat dan bagaimana hasil klasterisasi dapat dimanfaatkan untuk mendukung perumusan kebijakan pendidikan. Jawaban atas pertanyaan-pertanyaan ini tidak hanya akan memperkaya literatur akademik, tetapi juga memberikan dampak nyata bagi perbaikan sistem pendidikan di Indonesia.

## METODE PENELITIAN

Data penelitian yang digunakan dalam studi ini merupakan data sekunder. Data diambil dari Neraca Pendidikan Daerah (NPD) Provinsi Jawa Barat tahun 2019 yang dikeluarkan oleh Pusat Analisis dan Sinkronisasi Kebijakan, Sekretariat Jenderal, Kementerian Pendidikan dan Kebudayaan Republik Indonesia. NPD merupakan platform informasi tentang potret kinerja pendidikan pada suatu daerah yang dapat diakses oleh seluruh pemangku kepentingan. Adapun metode yang digunakan adalah analisis klustering dengan metode K-Means Clustering untuk mengelompokkan data berdasarkan kesamaan atau karakteristik tertentu tanpa memerlukan label atau target yang sudah ditentukan sebelumnya.

### Variabel

Penelitian ini memanfaatkan data yang diperoleh dari 27 kabupaten dan kota di wilayah Provinsi Jawa Barat. Data tersebut mencakup informasi mengenai kualifikasi guru, guru tersertifikasi serta distribusi atau pemerataan guru di tingkat Sekolah Menengah Pertama (SMP). Adapun penjelasan masing-masing variabel terangkum dalam Tabel 1 berikut.

**Tabel 1. Variabel Penelitian**

Nama variabel	Simbol	Keterangan
Kualifikasi guru	$X_1$	Persentase guru SMP yang sudah memenuhi kualifikasi S1/D4 pada tahun pelajaran 2018/2019.
Guru tersertifikasi	$X_2$	Persentase guru yang sudah tersertifikasi. Sertifikat pendidik diberikan kepada guru yang telah memenuhi standar profesional guru.
Indeks pemerataan guru PNS	$X_3$	Pemerataan sempurna guru dapat terjadi apabila ketersediaan guru PNS sama dengan kebutuhan guru pada semua sekolah di satu wilayah yang ditunjukkan oleh IPG dengan angka 0. Namun apabila terdapat ketimpangan penugasan yang menyebabkan adanya kekurangan atau kelebihan guru, IPG akan bergerak dari angka 0 menuju angka 1 (satu). Ketidakmerataan sempurna guru dicerminkan oleh IPG dengan angka 1 (satu).

### K-Means Clustering

Analisis klustering adalah metode statistik yang digunakan untuk mengelompokkan data ke dalam kelompok-kelompok (klaster) berdasarkan kemiripan atau karakteristik tertentu. Berdasarkan (Zhafar et al., 2023), data clustering merupakan proses untuk mengenali kelompok-kelompok atau klaster dalam data multidimensional dengan menggunakan berbagai ukuran kesamaan tertentu. (Al Azies & Rositawati, 2021) menyatakan analisis klaster terdiri dari dua jenis, yaitu klasterisasi hierarkis dan non-hierarkis. Pada klasterisasi hierarkis, jumlah kelompok yang akan terbentuk belum diketahui sebelumnya. Sementara itu, dalam klasterisasi non-hierarkis, jumlah kelompok yang akan dibentuk sudah ditentukan sebelumnya. Penelitian ini menggunakan salah satu metode klasterisasi non-hierarkis, yaitu K-Means Clustering.

K-Means Clustering adalah salah satu algoritma analisis clustering yang paling populer. Algoritma ini membagi dataset menjadi sejumlah klaster ( $k$ ) yang telah ditentukan sebelumnya. (Ma'ali et al., 2022) menyatakan algoritma k-means adalah salah satu metode *unsupervised machine learning*

yang bertujuan membagi data observasi ke dalam sejumlah kluster sebanyak  $k$ . Algoritma ini memerlukan inisialisasi awal berupa nilai centroid secara acak untuk setiap kluster. Pendekatan *unsupervised learning* ini tidak ada label atau kategori yang sudah ditentukan sebelumnya, sehingga algoritma harus menemukan pola tersembunyi dalam data secara mandiri. Setiap kluster direpresentasikan oleh sebuah centroid, yaitu titik pusat yang dihitung berdasarkan rata-rata posisi data dalam kluster tersebut. Tujuan utama dari K-Means Clustering adalah meminimalkan variasi dalam kluster dan memaksimalkan jarak antar kluster, sehingga menghasilkan pengelompokan yang optimal dan terstruktur.

Algoritma dari K-Means clustering adalah:

1. Input data.

- a. Menginput dataset  $X = \{x_1, x_2, x_3, \dots, x_n\}$  dimana setiap  $x_i$  merupakan data multidimensional dalam ruang dimensi-d.
- b. Menentukan jumlah kluster yang diinginkan (nilai  $k$ ).

2. Inisialisasi.

Menentukan secara acak  $k$  centroid awal, dinyatakan dalam  $C = \{\mu_1, \mu_2, \mu_3, \dots, \mu_k\}$  dari dataset. Berdasarkan (Ma'ali et al., 2022), pemilihan centroid awal dilakukan secara acak dari objek yang tersedia sesuai dengan jumlah kluster.

3. Iterasi.

a. Langkah 1: Penentuan kluster.

Untuk setiap titik data  $x_i$  dalam dataset, dihitung jaraknya ke semua centroid. (James et al., 2013) menyatakan jarak yang biasa digunakan adalah jarak euclidian, yaitu:

$$d(x_i, \mu_j) = \sqrt{\sum_{l=1}^d (x_{il} - \mu_{jl})^2} \quad (1)$$

Setiap data  $x_i$  dikelompokkan ke kluster dengan centroid terdekat:

$$C_i = \arg \min_{j \in \{1, 2, \dots, k\}} d(x_i, \mu_j) \quad (2)$$

b. Langkah 2: Pembaruan centroid.

Setelah semua data dikelompokkan, dihitung ulang posisi centroid setiap kluster  $j$  sebagai rata-rata posisi semua data dalam kluster tersebut:

$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \quad (3)$$

Di mana  $|C_j|$  adalah banyaknya data dalam kluster  $j$ .

Langkah 1 dan 2 diulang hingga tidak ada perubahan anggota kluster.

4. Output.

Diperoleh hasil sebagai berikut:

- a. Centroid akhir  $C = \{\mu_1, \mu_2, \mu_3, \dots, \mu_k\}$ .
- b. Kluster  $C_j$  yang berisi anggota-anggota data untuk setiap  $j = 1, 2, \dots, k$ .

Pemilihan  $k$  terbaik pada penelitian ini menggunakan metode koefisien Silhouette. Berdasarkan (Pramessti et al., 2017), koefisien Silhouette adalah metode yang digunakan untuk mengevaluasi kualitas dan kekuatan suatu kluster. Metode ini menggabungkan dua pendekatan, yaitu cohesi dan

separation. Pendekatan cohesion digunakan untuk mengukur sejauh mana hubungan antara objek dalam kluster yang sama, sementara separation digunakan untuk mengukur sejauh mana suatu kluster terpisah dari kluster lainnya. Nilai koefisien Silhouette berada dalam rentang antara -1 hingga 1. Semakin mendekati nilai 1, semakin baik kualitas pengelompokan data dalam satu kluster. Sebaliknya, jika nilainya mendekati -1, maka pengelompokan data dalam kluster tersebut semakin buruk.

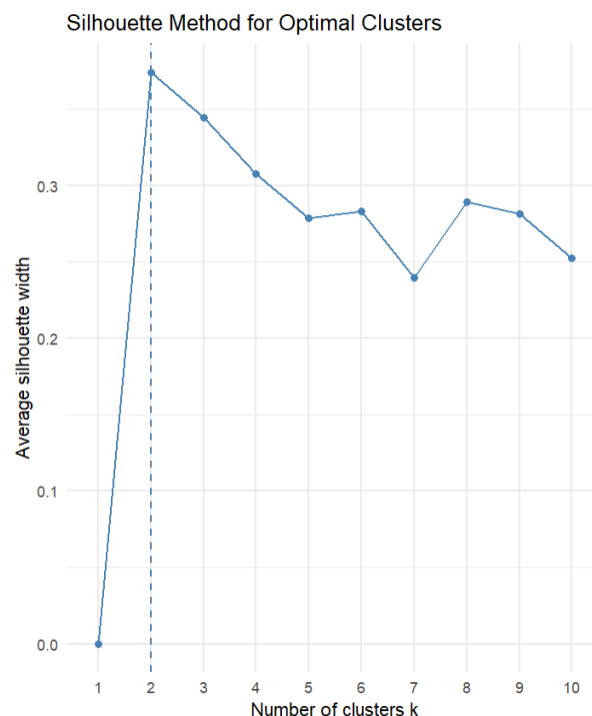
## HASIL DAN PEMBAHASAN

Data yang digunakan adalah data kualifikasi guru, guru tersertifikasi, dan indeks pemerataan guru PNS dari 27 kabupaten dan kota di Provinsi Jawa Barat yang diambil dari Neraca Pendidikan Daerah (NPD) tahun 2019. Data ini dikeluarkan oleh Pusat Analisis dan Sinkronisasi Kebijakan, Sekretariat Jenderal, Kementerian Pendidikan dan Kebudayaan Republik Indonesia.

**Tabel 2.** Deskripsi Data

Statistik deskriptif	$X_1$	$X_2$	$X_3$
Rata-rata	95,63	50,56	0,33
Standar deviasi	1,47	11,43	0,15
Nilai minimal	92,70	29,70	0,11
Nilai maksimal	98,10	75,50	0,65

Selanjutnya dilakukan analisis klustering pada dataset menggunakan bantuan software R. Untuk menentukan nilai  $k$ , dihitung koefisien Silhouette dan diperoleh hasil seperti pada gambar berikut:



**Gambar 1.** Koefisien *Silhouette*

Berdasarkan analisis menggunakan metode Silhouette, diperoleh bahwa nilai rata-rata koefisien Silhouette tertinggi dicapai pada jumlah kluster ( $k$ ) sebesar 2. Hal ini menunjukkan bahwa data memiliki struktur kluster yang paling optimal ketika dibagi menjadi dua kelompok. Dengan  $k = 2$ , jarak antar data dalam kluster (cohesion) menjadi paling erat, sementara jarak antar kluster (separation)

cukup jauh, mengindikasikan bahwa pembagian klaster ini memberikan kualitas pengelompokan yang terbaik dibandingkan dengan jumlah klaster lainnya.

Tahap selanjutnya adalah melakukan proses klasterisasi menggunakan algoritma K-Means pada dataset yang telah tersedia. Proses ini membagi data ke dalam klaster-klaster berdasarkan pola kemiripan antar data. Setelah proses klasterisasi selesai, hasil yang diperoleh menunjukkan pembagian data ke dalam dua klaster sesuai dengan nilai  $k$  yang telah ditentukan sebelumnya. Hasil lengkap dari klasterisasi K-Means ditampilkan sebagai berikut.

**Tabel 3.** Komposisi Klaster

Kabupaten/Kota	Klaster
Bogor	2
Sukabumi	2
Cianjur	2
Kabupaten Bandung	2
Sumedang	1
Garut	2
Tasikmalaya	2
Ciamis	1
Kuningan	1
Majalengka	1
Cirebon	1
Indramayu	2
Subang	2
Purwakarta	2
Karawang	2
Bekasi	2
Kabupaten Bandung Barat	2
Pangandaran	2
Kota Bandung	1
Kota Bogor	1
Kota Sukabumi	1
Kota Cirebon	1
Kota Bekasi	2
Kota Depok	2
Kota Cimahi	1
Kota Tasikmalaya	1
Kota Banjar	1

Hasil klasterisasi menunjukkan bahwa dari seluruh data yang dianalisis, sebanyak 12 Kabupaten/Kota tergolong ke dalam Klaster 1, sementara 15 Kabupaten/Kota lainnya tergabung dalam Klaster 2. Pembagian ini mengindikasikan adanya pola tertentu dalam karakteristik data yang memisahkan kedua klaster. Kabupaten/Kota yang masuk ke dalam Klaster 1 memiliki kesamaan atau kedekatan dalam variabel-variabel tertentu yang membedakannya dari Klaster 2, yang memiliki pola atau karakteristik yang berbeda. Jumlah ini menunjukkan bahwa kedua klaster memiliki distribusi data yang cukup seimbang, meskipun terdapat perbedaan kecil dalam jumlah anggota pada masing-masing klaster.

Pusat (center) dari setiap klaster yang dihasilkan melalui proses klasterisasi dirangkum dalam Tabel 4 berikut.

**Tabel 4.** Center dari Klaster

Klaster	$X_1$	$X_2$	$X_3$
1	0.4487388	0.8910606	-0.8216932
2	-0.3589911	-0.7128485	0.6573546

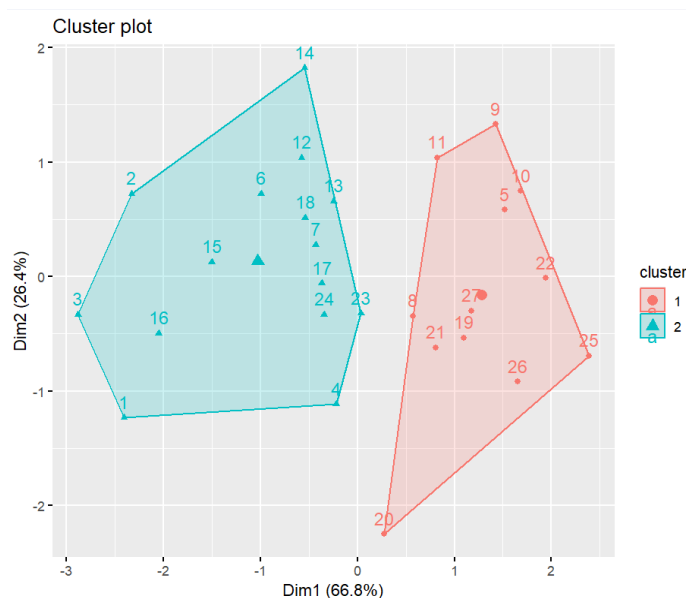
Center ini merepresentasikan nilai rata-rata dari setiap variabel untuk anggota kluster terkait, sehingga dapat digunakan sebagai gambaran umum karakteristik utama masing-masing kluster. Merujuk pada penjelasan (Talakua et al., 2017), karena data melalui proses standardisasi maka karakteristik data akan memiliki rata-rata 0 dan standar deviasi 1. Dalam konteks ini, center dengan nilai negatif menunjukkan bahwa data berada di bawah rata-rata keseluruhan data (lebih kecil dari rata-rata), sementara nilai positif mengindikasikan bahwa data berada di atas rata-rata keseluruhan data (lebih besar dari rata-rata).

Berdasarkan Tabel 4, terdapat perbedaan mencolok dalam karakteristik masing-masing kluster. Pada Klaster 1, variabel  $X_1$  dan  $X_2$  memiliki nilai di atas rata-rata, sementara  $X_3$  berada di bawah rata-rata. Sebaliknya, Klaster 2 menunjukkan pola yang berlawanan, di mana nilai  $X_1$  dan  $X_2$  berada di bawah rata-rata, sedangkan  $X_3$  berada di atas rata-rata. Pola ini mencerminkan distribusi data yang unik untuk masing-masing kluster, yang membantu dalam mengidentifikasi karakteristik dominan dalam setiap kelompok.

Pada Klaster 1, variabel  $X_1$  (persentase guru SMP yang telah memenuhi kualifikasi S1/D4 pada tahun pelajaran 2018/2019) dan  $X_2$  (persentase guru yang telah tersertifikasi) menunjukkan nilai yang lebih tinggi dibandingkan rata-rata keseluruhan. Hal ini mengindikasikan bahwa Klaster 1 didominasi oleh wilayah yang memiliki kualitas guru yang lebih baik. Namun, variabel  $X_3$  (indeks pemerataan guru PNS) berada di bawah rata-rata, yang menunjukkan distribusi guru yang lebih merata di wilayah ini, mengindikasikan pengelolaan tenaga pendidik yang lebih baik.

Sebaliknya, Klaster 2 menunjukkan karakteristik yang berbeda. Variabel  $X_1$  dan  $X_2$ , yang mengukur tingkat kualifikasi dan sertifikasi guru, berada di bawah rata-rata, mengindikasikan kualitas tenaga pengajar yang relatif lebih rendah di wilayah ini. Di sisi lain, variabel  $X_3$  berada di atas rata-rata, yang berarti pemerataan guru PNS di wilayah ini kurang optimal. Hal ini menunjukkan bahwa Klaster 2 mencakup wilayah-wilayah yang mungkin menghadapi tantangan dalam pemerataan dan kualitas sumber daya pendidikannya. Analisis ini memberikan gambaran yang jelas mengenai perbedaan karakteristik antara kedua kluster.

Hasil pengelompokan Kabupaten/Kota di Provinsi Jawa Barat divisualisasikan melalui diagram yang ditampilkan pada gambar berikut.



**Gambar 2.** Hasil Klastering

## SIMPULAN

Berdasarkan hasil analisis dengan K-Means Clustering, terdapat dua kluster utama dengan karakteristik yang mencolok di Provinsi Jawa Barat. Kluster 1 terdiri dari 12 Kabupaten/Kota, yaitu Sumedang, Ciamis, Kuningan, Majalengka, Cirebon, Kota Bandung, Kota Bogor, Kota Sukabumi, Kota Cirebon, Kota Cimahi, Kota Tasikmalaya, dan Kota Banjar. Pada kluster ini, variabel  $X_1$  (persentase guru SMP yang telah memenuhi kualifikasi S1/D4) dan  $X_2$  (persentase guru yang telah tersertifikasi) memiliki nilai di atas rata-rata, menunjukkan kualitas guru yang cukup baik dibandingkan rata-rata keseluruhan. Namun, variabel  $X_3$  (indeks pemerataan guru PNS) berada di bawah rata-rata, yang mengindikasikan distribusi guru yang lebih merata di wilayah-wilayah ini.

Sebaliknya, Kluster 2 mencakup 15 Kabupaten/Kota, yaitu Bogor, Sukabumi, Cianjur, Kabupaten Bandung, Garut, Tasikmalaya, Indramayu, Subang, Purwakarta, Karawang, Bekasi, Kabupaten Bandung Barat, Pangandaran, Kota Bekasi, dan Kota Depok. Wilayah-wilayah dalam kluster ini menunjukkan kekurangan pada semua variabel. Nilai  $X_1$  dan  $X_2$  berada di bawah rata-rata, mengindikasikan kualitas tenaga pendidik yang lebih rendah. Selain itu, variabel  $X_3$  berada di atas rata-rata, yang menunjukkan pemerataan guru PNS di wilayah ini kurang optimal, sehingga masih terdapat ketimpangan dalam distribusi tenaga pendidik.

Secara keseluruhan, hasil klusterisasi ini memberikan wawasan yang mendalam mengenai pola distribusi dan kualitas tenaga pendidik di Jawa Barat. Wilayah dalam Kluster 1 menunjukkan kualitas guru yang cukup baik dengan distribusi yang lebih merata, sedangkan Kluster 2 menghadapi tantangan dalam semua aspek, baik kualitas maupun pemerataan guru. Penelitian selanjutnya dapat mengeksplorasi metode clustering lain seperti hierarchical clustering untuk perbandingan hasil, serta mempertimbangkan penggunaan data primer seperti survei atau wawancara guna melengkapi analisis.

## REFERENSI

- Al Azies, H., & Rositawati, A. F. D. (2021). Mapping of the Reading Literacy Activity Index in East Java Province, Indonesia: an Unsupervised Learning Approach. *Proceedings of The International Conference on Data Science and Official Statistics*, 1, 211–223. <https://doi.org/10.34123/icdsos.v2021i1.128>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning - with Applications in R. In *Book*. Springer.
- Ma'ali, A. A., Girinoto, Ghiffari, M. N., & Hadiprakoso, R. B. (2022). Analisis Log Web Server dengan Pendekatan Algoritme K-Means Clustering dan Feature Importance. *Info Kripto*, 16(3), 119–123. <https://doi.org/10.56706/ik.v16i3.60>
- Pramesti, D. F., Furqon, M. T., & Dewi, C. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data otensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(9), 723–732. <https://doi.org/10.1109/EUMC.2008.4751704>
- Putri, M. A., & Abdullah, S. (2021). Clustering analysis of senior high school in West Java based on educational facilities. *Journal of Physics: Conference Series*, 2. <https://doi.org/10.1088/1742-6596/1725/1/012032>
- Suputra, I., Candiasa, I., & Suryawan, I. (2021). Klusterisasi Hasil Ujian Nasional SMA/MA dengan Algoritma K-Means. *Wahana Matematika Dan Sains: Jurnal Matematika, Sains, Dan*

*Pembelajarannya*, 15(1), 22–30. <https://ejournal.undiksha.ac.id/index.php/JPM/article/view/25380>

Talakua, M. W., Leleury, Z. A., & Taluta, A. W. (2017). Analisis Cluster Dengan Menggunakan Metode K-Means Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 11(2), 119–128. <https://doi.org/10.30598/barekengvol11iss2pp119-128>

Wibowo, A. E., & Habanabakize, T. (2022). K-Means Clustering untuk Klasifikasi Standar Kualifikasi Pendidikan dan Pengalaman Kerja Guru SMK di Indonesia. *Jurnal Dinamika Vokasional Teknik Mesin*, 7(2), 152–163. <https://doi.org/10.21831/dinamika.v7i2.53848>

Zhafar, M. N., Usman, K., & Akhyar, F. (2023). Penerapan Metode Clustering Dengan Algoritma K-Means Untuk Analisa Persebaran Varian Covid-19 ( Studi Kasus Kelurahan Antapani Kidul ). *E-Proceeding of Engineering*, 10(5), 4257–4264. <http://jurnal.mdp.ac.id>

