

Pengelompokan Kabupaten/Kota di Jawa Barat Berdasarkan Tingkat Kesejahteraan Masyarakat Menggunakan K-Means Cluster

Nabila Oktaviani ^{1*}, Achmad Fauzan ¹, Galuh Widyastuti ²

¹ Program Studi Statistika, Universitas Islam Indonesia, Jl. Kaliurang KM 14,5, Kabupaten Sleman Daerah Istimewa Yogyakarta, 55584, Indonesia

² Badan Pusat Statistika Kabupaten Kulonprogo

*corresponding author: nabila.oktaviani@students.uii.ac.id



P-ISSN: 2986-4178

E-ISSN: 2988-4004

Riwayat Artikel

Dikirim: 04 April 2024

Direvisi: 20 Juni 2024

Diterima: 26 Juni 2024

ABSTRAK

Tingkat kesejahteraan suatu bangsa dikatakan berhasil jika kehidupan masyarakatnya terpenuhi dengan baik. Penelitian ini bertujuan untuk mengidentifikasi tingkat kesejahteraan masyarakat di Provinsi Jawa Barat dengan metode pengelompokan K-Means berdasarkan indikator kesejahteraan rakyat. Hasil pengelompokan menunjukkan adanya dua kluster tingkat kesejahteraan: tinggi dan rendah, serta penggunaan tiga kluster untuk hasil yang lebih rinci. Kabupaten/kota di Jawa Barat dibagi menjadi kluster pertama dengan kesejahteraan tinggi (3 kabupaten/kota), kluster kedua dengan tingkat kesejahteraan sedang (18 kabupaten/kota), dan kluster ketiga dengan kesejahteraan rendah (6 kabupaten/kota). Penggunaan tiga kluster lebih baik karena memberikan pemahaman yang lebih terperinci mengenai distribusi kesejahteraan masyarakat. Hasil klasifikasi menunjukkan terdapat ketimpangan kesejahteraan antar wilayah di Jawa Barat, memungkinkan identifikasi target pembangunan yang lebih efektif untuk mencapai tujuan pembangunan yang inklusif dan berkelanjutan.

Kata Kunci: Kesejahteraan Masyarakat, Jawa Barat, Algoritma K-Means.

ABSTRACT

Welfare is a universal aspiration desired by every individual and serves as a fundamental pillar of high hopes and aspirations in the struggle of the Indonesian nation. The level of welfare of a nation is one of the main indicators to assess the success of development in a country. This study aims to identify the level of welfare of society in West Java Province using the K-Means clustering method based on indicators of people's welfare. The results of the clustering show the existence of two clusters of welfare levels: high and low, and the use of three clusters for more detailed results. Districts/cities in West Java are divided into the first cluster with high welfare (3 districts/cities), the second cluster with moderate welfare levels (18 districts/cities), and the third cluster with low welfare (6 districts/cities). In conclusion, there is an imbalance of welfare among regions in West Java, enabling the identification of more effective development targets to achieve inclusive and sustainable development goals.

Keywords: Welfare of Society, West Java, K-Means Algorithm.

1. Pendahuluan

Kesejahteraan merupakan aspirasi universal yang diinginkan oleh setiap individu. Hal ini menjadi pilar utama harapan dan cita-cita yang tinggi dalam perjuangan bangsa Indonesia [1]. Tingkat kesejahteraan suatu bangsa menjadi salah satu indikator utama untuk menilai keberhasilan pembangunan di suatu negara, konsep kesejahteraan mencakup aspek keamanan, kesejahteraan, dan kemakmuran yang tercermin dalam kualitas hidup. Secara umum, kesejahteraan dapat dijelaskan sebagai keadaan di mana seluruh warga negara memperoleh kecukupan dalam segala aspek kebutuhan, baik yang bersifat materil maupun spiritual. Definisi ini diperkuat oleh Faisal Basri (1993), yang menjelaskan bahwa kesejahteraan merupakan kondisi di mana individu-individu dalam suatu masyarakat dapat memenuhi kebutuhan hidup Masyarakat dengan baik, tidak hanya secara materi tetapi juga dalam hal-hal yang bersifat spiritual dan emosional.

Pencapaian kesejahteraan masyarakat menjadi tujuan utama dalam setiap upaya Pembangunan ekonomi. Di Indonesia, pencapaian kesejahteraan juga menjadi salah satu tujuan negara, sebagaimana dijelaskan dalam pembukaan UUD 1945 alenia ke-empat, yang menegaskan komitmen untuk melindungi segenap bangsa Indonesia dan seluruh tumpah darah Indonesia serta untuk memajukan kesejahteraan umum, termasuk di Tingkat regional seperti Jawa Barat [2].

Jawa Barat merupakan salah satu provinsi di Indonesia yang terletak di bagian barat Pulau Jawa. Provinsi ini memiliki peran penting dalam Sejarah, budaya, dan ekonomi Indonesia. Dengan luas wilayah 35.377,76 km², Jawa Barat merupakan salah satu provinsi terpadat di Indonesia. Secara ekonomi, Jawa Barat memiliki kontribusi yang signifikan terhadap perekonomian nasional. Menurut data dari Badan Pusat Statistik Provinsi Jawa Barat, tingkat kemiskinan pada tahun 2022 mencapai 8,06%, menunjukkan penurunan dari tahun sebelumnya, yaitu 8,40% pada tahun 2021. Meskipun permasalahan kemiskinan menjadi perhatian utama pemerintah, namun Tingkat pengangguran di Jawa Barat masih tetap tinggi. Berdasarkan data tahun 2022, total Jumlah Pengangguran Terbuka adalah 2.125.606, turun -14.33% (opendata.jabarprov.go.id). Meskipun tingkat kemiskinan dan pengangguran menurun, pertumbuhan kesempatan kerja tidak mampu mengimbangi laju pertumbuhan angkatan kerja setiap tahunnya. Oleh karena itu, diperlukan perubahan strategi dan inovasi baru dalam upaya penciptaan lapangan kerja yang lebih efektif. Dalam menjalankan proses perubahan dan penting untuk melakukan identifikasi berdasarkan karakteristik tingkat kesejahteraan masyarakat di setiap wilayah atau daerah. Hal ini bertujuan agar kebijakan dan strategi pembangunan dapat disesuaikan dengan kebutuhan yang spesifik dan mencapai sasaran yang tepat. Proses Pembangunan dapat berhasil dan mencapai sasaran yang tepat, jika dilakukan pengidentifikasian wilayah target yang tepat [3]. Oleh karena itu perlu dilakukannya pengelompokan karakteristik kabupaten/kota yang ada di provinsi Jawa Barat berdasarkan indikator kesejahteraan rakyat.

Studi terkait pengelompokan wilayah berdasarkan indikator kesejahteraan telah banyak dilakukan. Misalnya, penelitian yang dilakukan oleh Niken Aushaf Nabilah (2024) tentang pengelompokan provinsi di Indonesia berdasarkan indikator kesejahteraan masyarakat dengan algoritma k-means. Penelitian lainnya dilakukan oleh Ahmad Dedi Hartono, Deni Afirianto, dan Ilham Saifudin (2022) tentang implementasi algoritma K-Means dengan *Particle Swarm Optimization* untuk pengelompokan tingkat kesejahteraan di Provinsi Jawa Timur. Secara keseluruhan penelitian penelitian ini memberikan kontribusi signifikan dalam memahami distribusi kesejahteraan di berbagai wilayah dan mendukung perumusan kebijakan yang lebih berkeadilan dan efektif.

Pengelompokan karakteristik kabupaten/kota yang ada di provinsi Jawa Barat menggunakan metode statistika non hierarkis, yaitu algoritma K-Means. Dalam algoritma K-Means, data dikelompokkan menjadi k kelompok berdasarkan kesamaan karakteristik tertentu. Melalui analisis K-Means, pengelompokan Kabupaten/kota berdasarkan indikator kesejahteraan dapat dilakukan, serta untuk mengidentifikasi daerah-daerah di

Jawa Barat yang memerlukan prioritas bantuan dalam rangka meningkatkan efektivitas program Pembangunan pemerintah.

2. Metodologi Penelitian

2.1 Data dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik Provinsi D.I Yogyakarta (2022) [4]. Data tersebut berupa data indikator kesejahteraan masyarakat di Provinsi Jawa Barat.

2.2 Definisi Variabel

Variabel yang digunakan dalam penelitian ini dapat dijelaskan dalam tabel berikut.

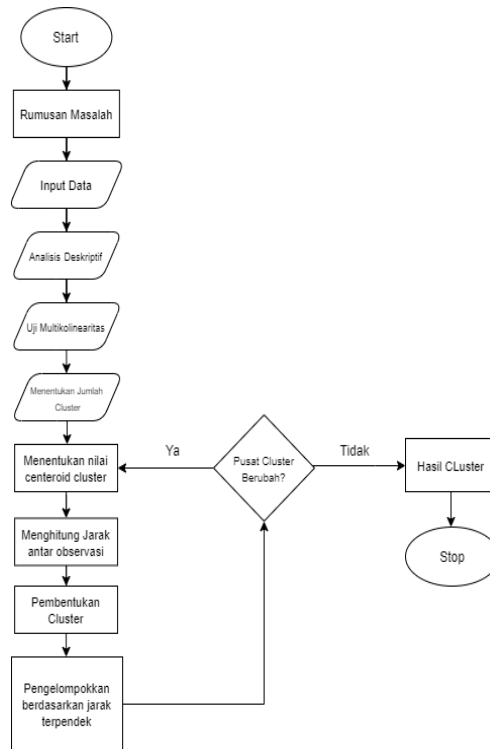
Tabel 1 Variabel dan Definisi Operasional Variabel Penelitian

Variabel	Definisi
Laju Pertumbuhan Penduduk (LPP)	Variabel Laju Pertumbuhan Penduduk berisi ukuran yang digunakan untuk menggambarkan perubahan jumlah penduduk suatu wilayah dalam suatu periode.
Pengeluaran Perkapita (PP)	Variabel Pengeluaran Perkapita mengacu pada jumlah total pengeluaran yang dibelanjakan atau dihabiskan dalam suatu wilayah.
Kepadatan Penduduk (KP)	Variabel Kepadatan Penduduk mengacu pada jumlah rata-rata individu yang tinggal di suatu wilayah.
Angkatan Kerja (AK)	Variabel Angkatan Kerja mengacu pada jumlah individu dalam populasi yang aktif mencari pekerjaan atau bekerja pada suatu titik waktu tertentu.

2.3 Metode Penelitian

Penelitian ini menggunakan metode K-Means *Clustering* dengan bantuan *software RStudio* untuk menganalisis data dengan menggunakan 2 kluster dan 3 kluster. Nilai k sebanyak 2 kluster yaitu tinggi dan rendah dan nilai k sebanyak 3 kluster yaitu tinggi, sedang, dan rendah. Nilai k tersebut mewakili tingkat kesejahteraan Masyarakat di Provinsi Jawa Barat. Metode ini digunakan untuk membagi wilayah kabupaten/kota ke dalam kelompok yang berbeda berdasarkan karakteristik yang saling mirip. Dengan menerapkan K-Means *Clustering* dengan 2 kluster, penelitian ini bertujuan untuk mengidentifikasi dua kelompok yang memiliki pola atau karakteristik yang berbeda. Sedangkan dengan menggunakan 3 kluster, penelitian ini akan mengidentifikasi tiga kelompok dengan pola yang berbeda pula untuk mendapatkan pemahaman yang lebih komprehensif tentang pola yang ada dalam dataset. Berikut penggambaran algoritma K-Means *Clustering* dengan menggunakan diagram alir pada **Gambar 1**. Adapun algoritmanya adalah sebagai berikut.

1. Merumuskan masalah yang diangkat dalam penelitian.
2. Menginput data
3. Melakukan analisis deskriptif untuk melihat gambaran umum dari data jumlah siswa dan jumlah guru di provinsi NTB.
4. Melakukan uji multikolinearitas, pengecekan multikolinearitas ini dilakukan karena jika terdapat multikolinearitas maka hal tersebut merupakan suatu pelanggaran dalam analisis cluster, karena dapat mempengaruhi hasil *cluster*. Uji multikolinearitas dilakukan dengan menggunakan nilai VIF.
5. Menentukan jumlah *cluster k* dengan metode *Silhouette*.
6. Menentukan nilai *centeroid cluster*
7. Menghitung jarak antar observasi dengan menggunakan jarak *Euclidean*.
8. Melakukan pembentukan *cluster*.
9. Melakukan pengelompokan berdasarkan jarak terpendek.
10. Interpretasi hasil *cluster*.



Gambar 1 Langkah-Langkah Pengklasteran dengan Menggunakan Metode *K-Means*

2.3.1 Statistika Deskriptif

Statistika adalah ilmu yang melibatkan proses pengeumpulan data, pengolahan data, presentasi data, analisis data, inferensi, dan pengambilan keputusan yang didasarkan pada fakta-fakta yang ada. Statistika adalah ilmu pengetahuan yang berhubungan dengan metode-metode pengumpulan data, pengolahan data, penyajian data, penganalisisan data, penarikan kesimpulan serta membuat keputusan yang cukup beralasan berdasarkan fakta yang ada [5]. Statistik deskriptif fokus pada pemahaman dan penyajian informasi terkait data atau keadaan tertentu. Dalam statistik deskriptif, data disajikan dengan menguraikan atau memberikan keterangan mengenai data tersebut. Salah satu cara yang umum digunakan adalah melalui ukuran pemusatan data, seperti nilai maksimum, nilai minimum, dan nilai rata-rata atau mean pada setiap variabelnya [6].

2.3.2 Data Mining

Data *mining* adalah serangkaian metode yang digunakan untuk mengungkap nilai tambahan dari suatu kumpulan data dengan cara menemukan pengetahuan yang tidak dapat ditemukan secara manual. Dalam literatur data *mining*, terdapat beberapa teknik yang sering disebut pengelompokan, klasifikasi, jaringan saraf, algoritma genetika, dan lainnya [7]. Perangkat lunak penambangan data merupakan salah satu alat analisis yang digunakan untuk menganalisis data. Perangkat ini memungkinkan pengguna untuk menganalisis data dari berbagai dimensi atau sudut, mengkategorikan data, dan merangkum hubungan yang teridentifikasi, secara teknis, data mining sering dikenal sebagai knowledge discovery from data (KDD), sementara yang lain melihat data mining hanya sebagai Langkah penting dalam proses penemuan pengetahuan. Proses penemuan pengetahuan terdiri dari tahapan-tahapan seperti pembersih data, integrasi data, seleksi data, transformasi data, penambangan data, evaluasi pola, dan presentasi pengetahuan [8].

2.3.3 Analisis Clustering

Analisis *Cluster* adalah metode umum yang digunakan untuk pengelompokan data, dimana data dalam suatu himpunan dibagi ke dalam beberapa kelompok berdasarkan tingkat kesamaannya, dengan kesamaan data dalam suatu kelompok lebih besar dibandingkan dengan data dalam kelompok lain. [9]

Analisis *Cluster* mempunyai tujuan untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *cluster* mengklasifikasi objek sehingga objek-objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* sama. Dalam analisis *cluster*, terdapat dua metode pengklasteran yang umum digunakan, yaitu metode *hierarchical* dan metode *non-hierarchical*. Metode *hierarchical* adalah pendekatan yang dimana data diklasterkan dengan cara mengukur jarak antara setiap pasangan objek, dan kemudian membentuk sebuah pohon yang disebut dendrogram. Sedangkan metode *non-hierarchical* dimulai dengan menentukan jumlah *cluster* yang diinginkan sebelumnya. Setelah jumlah *cluster* ditentukan, proses pengklasteran dilakukan. Dalam metode ini, objek-objek akan dikelompokkan berdasarkan kesamaan karakteristik atau jarak antara objek tersebut. [10]

2.3.4 Cluster Validation

Klasifikasi tanpa pengawasan perlu divalidasi untuk memastikan kualitas pengelompokan yang dihasilkan. Salah satu metode validasi yang digunakan adalah koefisien *silhouette*. Koefisien *silhouette* adalah metode validasi yang digunakan untuk mengukur seberapa baik sebuah objek ditempatkan dalam klasternya dalam konteks klasifikasi tanpa pengawasan. Matriks ini menggambarkan seberapa dekat suatu objek dengan objek lain dalam klaster yang sama, sekaligus mempertimbangkan seberapa jauh objek tersebut dari objek terdekatnya. Rentang nilai koefisien *silhouette* adalah -1 hingga 1, dimana nilai yang lebih tinggi menunjukkan bahwa objek tersebut lebih cocok dengan klasternya sendiri dibandingkan dengan klaster terdekatnya. Nilai satu menandakan bahwa objek tersebut ditempatkan dengan sangat baik dalam klasternya, sedangkan nilai negatif menunjukkan bahwa objek tersebut mungkin ditempatkan di klaster yang salah. Koefisien *silhouette* berguna untuk mengevaluasi kualitas pengelompokan yang dihasilkan, dan membantu untuk memahami seberapa baik objek-objek dikelompokkan berdasarkan kesamaan.

Nilai koefisien *silhouette* untuk suatu objek, misalnya berada dalam rentang [-1, 1]. Jika nilai *silhouette* objek mendekati 1, itu menunjukkan bahwa objek tersebut merupakan representasi cluster yang kuat (memiliki kohesi yang kuat) dan cukup jauh dari cluster tetangga (memiliki pemisahan yang kuat). Jika nilai *silhouette* objek mendekati 0, itu berarti objek tersebut berada di dekat batas keputusan antara dua cluster tetangga. Jika nilai *silhouette* objek negatif, itu menunjukkan bahwa objek mungkin telah salah diklasifikasikan ke dalam cluster yang tidak sesuai [11].

2.3.5 Uji Multikolinearitas

Uji multikolinearitas adalah suatu metode yang digunakan untuk mengevaluasi adanya korelasi yang signifikan antara variabel-variabel bebas dalam suatu model regresi linier berganda. Jika terdapat korelasi yang kuat antara variabel-variabel bebas tersebut, maka hubungan antara variabel bebas dan variabel terikat dapat menjadi terganggu. Uji multikolinearitas dilakukan dengan memperhatikan nilai *Tolerance* dan *VIF* (*Variance Inflation Factor*), serta melihat tingkat korelasi antara variabel-variabel independen. Berikut adalah rumus untuk menghitung nilai *VIF* (*Variance Inflation Factor*):

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

Untuk menyimpulkan bahwa suatu model regresi bebas dari multikolinearitas, nilai *VIF* seharusnya tidak melebihi 10, dan nilai *tolerance* seharusnya tidak kurang dari 0,10. Dengan kata lain, jika model regresi memiliki nilai *VIF* yang rendah dan nilai *tolerance* yang tinggi, dapat dikatakan bahwa model tersebut bebas dari multikolinearitas [12].

2.3.6 K-Means Clustering

K-Means Clustering merupakan salah satu metode *non-hierarchical* yang berbasis partisi yang dapat mengelompokkan data berdasarkan titik pusat klaster (*centroid*) terdekat [13]. Metode ini mengelompokkan data ke dalam klaster dengan mempertimbangkan karakteristik yang serupa, sehingga data dengan karakteristik yang berbeda dikelompokkan dalam klaster yang berbeda. *K-Means clustering* bertujuan untuk mengurangi variasi antara

data yang berada dalam satu kluster, sambil memaksimalkan variasi data antara kluster yang berbeda [14].

K-Means *clustering* menggunakan parameter *input* k dan mengelompokkan sekumpulan n objek ke dalam k kluster. Tujuan utamanya adalah memastikan tingkat kemiripan antara anggota dalam satu kluster menjadi tinggi, sementara tingkat kemiripan dengan anggota kluster lainnya menjadi rendah. Untuk mengukur kemiripan anggota dengan kluster, digunakan konsep kedekatan objek terhadap nilai rata-rata (mean) pada *centroid* kluster [15]. Berikut ini adalah langkah-langkah dalam algoritma *K-Means*.

1. Menentukan jumlah *cluster* (k) pada *dataset*
2. Menentukan *centroid*, penentuan nilai *centroid* pada tahap awal ditentukan secara random, sedangkan pada tahap iterasi digunakan rumur seperti pada persamaan berikut ini:

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

Keterangan :

V_k = Centeroid pada *cluster* ke k

X_i = Data ke i

N_k = Banyaknya objek/jumlah data yang menjadi anggota *cluster* ke k

3. Menghitung jarak terdekat *centroid*. Jarak *centroid* yang digunakan adalah *Euclidean Distance*, dengan rumus seperti pada persamaan dibawah ini:

$$D_E = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (3)$$

Keterangan:

D_E = *Euclidean Distance*

i = Banyaknya Objek

(x, y) = Koordinat Objek

(s, t) = Koordinat *Centroid*

4. Alokasikan setiap data ke dalam *cluster* yang mempunyai jarak terdekat.
5. Lakukan kembali Langkah kedua dan ketiga sampai *centroid* stabil.

3. Hasil dan Pembahasan

3.1. Analisis Deskriptif

Analisis deskriptif dilakukan dengan tujuan untuk memperoleh pemahaman yang lebih mendalam tentang data yang ada. Dengan melakukan analisis deskriptif, peneliti dapat mengidentifikasi karakteristik dari data tersebut.

Tabel 2 Hasil Analisis Deskriptif

Variabel	Minimum	Mean	Maksimum
LPP	0.410	1.301	1.860
PP	16217	24496	40193
KP	383	3824	14776
AK	94831	501450	2897332

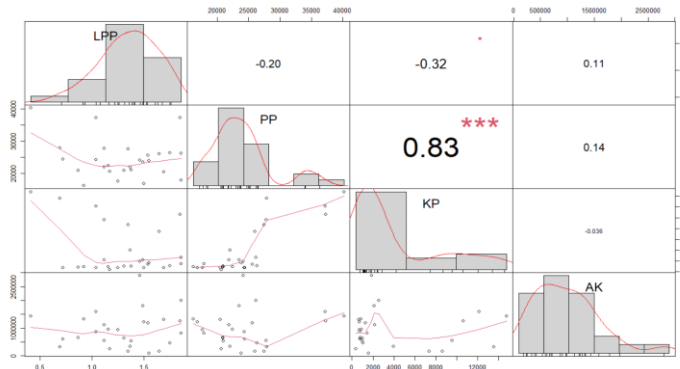
Berdasarkan hasil analisis deskriptif pada **Tabel 2**, dapat disimpulkan bahwa di Jawa Barat rata-rata laju pertumbuhan penduduk adalah 1,301%, dengan kisaran nilai antara 0,410% hingga 1.860%. Mayoritas kabupatennya menunjukkan laju pertumbuhan penduduk berkisar antara 1080% dan 1545%. Perihal pengeluaran perkapita, rata-ratanya mencapai 24.496 dengan minimum sekitar 16.217 dan maksimum sekitar 40.193. Mayoritas kabupaten menunjukkan pengeluaran perkapita antara 20.941 dan 26.272. Kepadatan penduduk rata-ratanya sekitar 3.824 orang/km², dengan rentang nilai antara 383 14.776. Mayoritas kabupaten menunjukkan kepadatan penduduk di bawah 6.682 orang/km². Selain itu, rata-rata Angkatan kerja adalah 947.340 orang, dengan nilai

minimum 94.831 dan maksimum 2.897.332. Mayoritas kabupaten memiliki jumlah agkatan kerja dibawah 1.286.322

3.2. Asumsi Analisis Cluster

3.2.1. Uji Multikolinearitas

Multikolinearitas adalah kondisi di mana terdapat korelasi atau hubungan yang kuat antara variabel bebas atau anantara beberapa variabel bebas dalam analisis regresi. Multikolinearitas mengindikasikan bahwa variabel bebas tidak saling independen satu sama lain [16]. Dalam analisis kluster penting untuk melakukan pengecekan multikolinearitas karena adanya multikolinearitas dapat mempengaruhi hasil kluster yang diperoleh. Multikolinearitas dianggap sebagai suatu pelanggaran dalam analisis kluster karena dapat mengaburkan pola sebenarnya dalam data [17].



Gambar 2 Korelasi antar Variabel

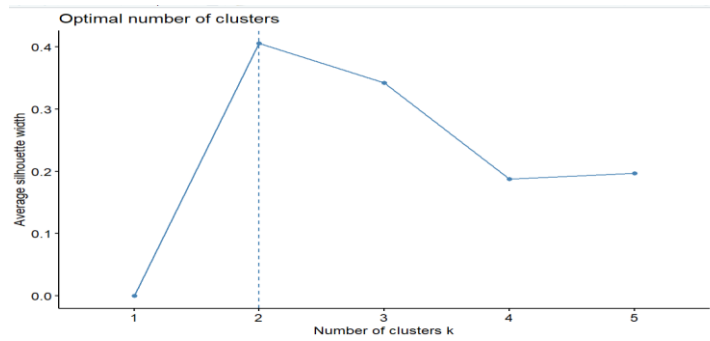
Berdasarkan **Gambar 2** dapat diamati bahwa korelasi antar variabel memiliki nilai di bawah 10 (<10), Sebagai contoh terdapat korelasi sebesar 0,83 antara variabel LPP dan variabel KP, korelasi tersebut lebih dari 50%. Korelasi yang tinggi ini menunjukkan adanya multikolinearitas yang kuat antara kedua variabel tersebut.

3.3. Analisis K-Means Clustering

3.3.1. Penentuan Jumlah Kluster

Peneliti menggunakan metode K-Means (*non-hierarchical*) untuk menentukan jumlah kluster yang optimal. Pada metode ini, peneliti memiliki kebebasan untuk menentukan jumlah kluster yang di inginkan. Sementara itu, pada metode hirarki, jumlah kluster ditentukan berdasarkan struktur dendrogram yang terbentuk. Penentuan Jumlah kluster yang optimal merupakan keputusan subyektif dan bergantung pada metode pengukuran kesamaan yang digunakan serta parameter-parameter yang digunakan untuk partisi data. Terdapat beberapa metode yang dapat digunakan untuk mengoptimalkan penentuan jumlah kluster dalam analisis kluster *non-hierarchical*, antara lain: metode *silhouette*, *elbow*, dan *gap-statistics*. Dalam penelitian ini, peneliti menggunakan metode *silhouette* untuk menentukan jumlah kluster yang optimal yang disajikan pada **Gambar 3**. Metode *silhouette* menggunakan rata-rata nilai *silhouette* untuk setiap data sebagai acuan dalam menentukan keoptimalan jumlah kluster. Semakin tinggi nilai rata-rata *silhouette*, maka semakin optimal jumlah kluster yang dipilih. Dengan menggunakan metode *silhouette* peneliti berharap dapat menentukan jumlah kluster yang memberikan hasil yang optimal dalam analisis kluster ini [17].

Berdasarkan **Gambar 3**, hasil penentuan kluster menggunakan metode *silhouette* menunjukkan bahwa nilai rata-rata *silhouette* tertinggi terjadi ketika jumlah kluster (k) adalah 2. Oleh karena itu, dalam penelitian ini, jumlah kluster yang digunakan adalah 2 yang menghasilkan dua kelompok yang berbeda yaitu terdapat kelompok rendah dan tinggi



Gambar 3 Metode *Silhouette*

Selain itu untuk memberikan informasi yang lebih rinci tentang variasi dalam data data, peneliti juga melakukan pengelompokan dengan jumlah kluster (k) sebanyak 3, yang menghasilkan tiga kelompok yang berbeda. Pendekatan ini memungkinkan peneliti untuk mengungkapkan perbedaan yang lebih spesifik dalam karakteristik dan pola antara kelompok-kelompok tersebut. Dengan menggunakan pendekatan dua atau tiga kluster, penelitian ini dapat memberikan pemahaman yang lebih mendalam tentang struktur data dan mengidentifikasi perbedaan yang lebih spesifik antara kelompok-kelompok yang terbentuk.

3.3.2. Menentukan Nilai Centeroid Kluster

Tujuan dalam penentuan nilai *centroid* kluster menggunakan algoritma K-Means adalah untuk mengelompokkan data menjadi kelompok-kelompok yang saling berdekatan berdasarkan kemiripan atau jarak antara data. Penentuan nilai *centroid* kluster dilakukan dengan meminimalkan variansi intra-kluster dan memaksimalkan variansi antara kluster.

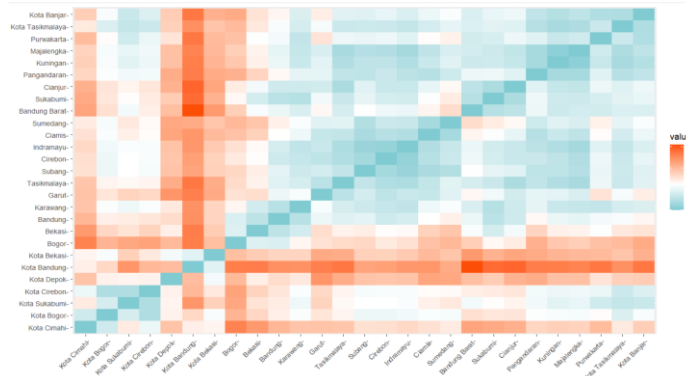
Tabel 3 Titik Pusat Kluster (*Centeroid*) 3 Kluster

Variabel	Cluster 1	Cluster 2	Cluster 3
LPP	0.80089752	0.64420904	-0.59307589
PP	0.01836009	-0.47928146	1.42866434
KP	-0.3595164	-0.5065024	1.6992655
AK	1.99172202	-0.30139088	-0.09168836

Pada **Tabel 3**, terdapat hasil titik pusat kluster (*centroid*) untuk setiap kluster dengan pembagian menjadi tiga kluster yaitu tinggi, sedang, dan rendah.

3.3.3. Menghitung Jarak antar Observasi

Dalam proses pengelompokan objek berdasarkan kesamaannya, salah satu langkah yang dilakukan adalah mengukur seberapa dekat atau jauhnya objek satu dengan objek lainnya. Peneliti menghitung jarak antar observasi dan mengvisualisasikannya dengan *heatmap* berikut.



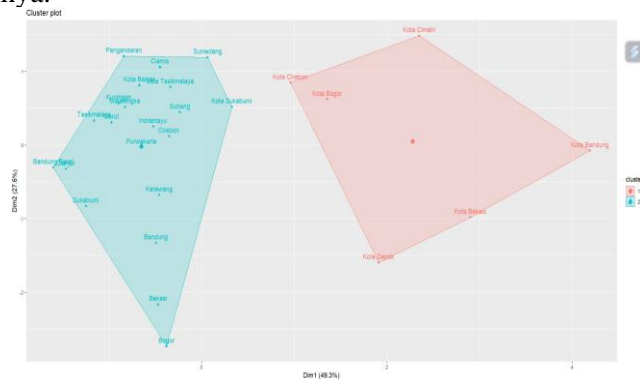
Gambar 4 *Heatmap* Jarak Antara Kabupaten

Berdasarkan **Gambar 4**, dapat ditarik kesimpulan bahwa terdapat pola hubungan antara kedekatan kabupeaten-kabupaten dalam hal jarak. Kabupaten-kabupaten yang memiliki kedekatan yang cukup jauh ditunjukkan oleh warna *orange* yang semakin gelap.

Sementara itu, kabupaten-kabupaten yang memiliki kedekatan yang lebih dekat ditunjukkan dengan warna biru. Adapun warna putih menandakan ukuran kedekatan berdasarkan jarak yang menengah. Pada heatmap tersebut dapat dilihat bahwa Kota Cirebon dan Kota Bogor ditunjukkan oleh warna biru, yang menunjukkan bahwa kedua kota tersebut memiliki kedekatan yang dekat dalam hal jarak. Di sisi lain, Kota Bogor dan Kota Cimahi ditunjukkan oleh warna orange pekat, yang berarti bahwa kedua kabupaten tersebut memiliki jarak yang cukup jauh.

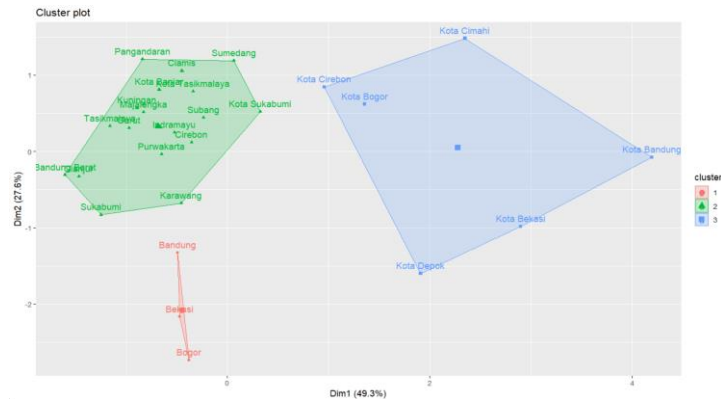
3.3.4. Pembentukan Klaster

Berdasarkan hasil pengelompokan dengan menggunakan klaster sebanyak 2 dan 3 klaster, dilakukan visualisasi hasil pengelompokannya . Berikut adalah hasil pengelompokannya.



Gambar 5 Visualisasi Hasil Pengelompokan 2 Klaster

Pada **Gambar 5**, terlihat plot hasil klaster yang menggunakan 2 klaster. Visualisasi tersebut menggunakan dua warna yang masing-masing mengidentifikasi klaster. Warna merah digunakan untuk mengidentifikasi klaster 1, warna hijau digunakan untuk mengidentifikasi klaster 2.



Gambar 6 Visualisasi Hasil Pengelompokan 3 Klaster

Pada **Gambar 6**, terlihat plot hasil klaster yang menggunakan 3 klaster. Visualisasi tersebut menggunakan tiga warna yang masing-masing mengidentifikasi klaster. Warna merah digunakan untuk mengidentifikasi klaster 1, warna hijau digunakan untuk mengidentifikasi klaster 2, dan warna biru mengidentifikasi klaster 3.

3.3.5. Hasil Pengelompokan

Dalam pengelompokan wilayah berdasarkan tingkat pendidikan dengan menggunakan metode K-Means dan data jumlah guru dan jumlah sekolah di Provinsi Nusa Tenggara Barat, dilakukan dengan menggunakan 2 klaster dan 3 klaster. Berdasarkan hasil visualisasi pengelompokan pada **Gambar 5** dan **6** diperoleh anggota-anggota dalam kelompok sebagai berikut.

Tabel 4 Hasil K-Means 3 Klaster

Pengelompokkan	Kabupaten	Jumlah Anggota
Kelompok 1	Kota Bandung, Kota Bekasi, Kota Bogor, Kota Cirebon, Kota Depok, Kota Cimahi	6
Kelompok 2	Pangandaran, Sumedang, Ciamis, Kota Banjar, Kota Tasikmalaya, Kota Sukabumi, Subang, Kuningan, Majalengka, Indramayu, Cirebon, Purwakarta, Sukabumi, Karawang, Bandung Barat, Cianjur, Sukabumi, Karawang, Bandung, Bekasi, Bogor	21

Berdasarkan **Tabel 4** hasil pengklasteran didapatkan 2 klaster dengan jumlah anggota pada klaster pertama 6 kabupaten/kota, pada klaster kedua 21 kabupaten/kota di Jawa Barat.

Tabel 5 Hasil K-Means 3 Klaster

Pengelompokkan	Kabupaten	Jumlah Anggota
Kelompok 1	Bandung, Bekasi, Bogor	3
Kelompok 2	Pangandaran, Sumedang, Ciamis, Kota Banjar, Kota Tasikmalaya, Kota Sukabumi, Subang, Kuningan, Majalengka, Indramayu, Cirebon, Purwakarta, Tasikmalaya, Garut, Sukabumi, Karawang, Bandung Barat, Cianjur.	18
Kelompok 3	Kota Cimahi, Kota Cirebon, Kota Bogor, Kota Bandung, Kota Bekasi, Kota Depok.	6

Berdasarkan **Tabel 5** hasil pengklasteran didapatkan 3 klaster dengan jumlah anggota pada klaster pertama 3 kabupaten/kota, pada klaster kedua 18 kabupaten/kota, dan klaster ke tiga 8 kabupaten/kota di Jawa Barat.

3.3.6. Profilisasi Data

Profilisasi data dilakukan untuk menggambarkan karakteristik dari masing-masing kelompok. Hal ini dilakukan dengan menggunakan rata-rata dari nilai pada setiap klaster.

Tabel 6 Profilisasi Data dengan K-Means 2 Klaster

Variabel	Cluster 1	Cluster 2
LPP	1.076667	1.364762
PP	32974.50	22073.19
KP	11551.667	1616.048
AK	887945.5	964309.6

Pada **Tabel 6**, dapat disimpulkan bahwa kabupaten/kota pada kluster 1 memiliki karakteristik laju pertumbuhan penduduk lebih rendah, namun dengan tingkat pengeluaran perkapita, kepadatan penduduk dan angkatan kerja yang tinggi. Sementara itu, kabupaten/kota pada kluster 2, menunjukkan laju pertumbuhan penduduk sedikit lebih tinggi dengan pengeluaran perkapita yang lebih rendah, serta kepadatan penduduk dan Angkatan kerja yang lebih rendah di dibandingkan kabupaten/kota pada kluster 1.

Tabel 7 Profilisasi Data dengan K-Means 3 Klaster

Variabel	Cluster 1	Cluster 2	Cluster 3
LPP	1.603333	1.076667	1.325000
PP	24604.67	32974.50	21651.28
KP	2189.000	11551.667	1520.556
AK	2237546.0	887945.5	752103.5

Pada **Tabel 7**, dapat dilihat bahwa pada kluster 1 menonjol dengan laju pertumbuhan penduduk (LPP) yang tinggi sekitar 1.60%, pengeluaran perkapita (PP) yang relatif tinggi sekitar 24,604.67, kepadatan penduduk (KP) yang cukup tinggi sekitar 2,189 orang per kilometer persegi, dan jumlah angkatan kerja (AK) yang signifikan sekitar 2,237,546 orang. Kluster ini mungkin mencakup daerah perkotaan yang padat penduduk dengan tingkat pengeluaran dan kegiatan ekonomi yang tinggi. Pada kluster 2 memiliki laju pertumbuhan penduduk (LPP) yang lebih rendah, sekitar 1.08%, namun memiliki pengeluaran perkapita (PP) tertinggi di antara kluster lainnya, sekitar 32,974.50. Kepadatan penduduk (KP) juga tinggi, mencapai 11,551.67 orang per kilometer persegi, sementara jumlah angkatan kerja (AK) lebih rendah dibandingkan dengan Kluster 1, sekitar 887,945.5 orang. Kluster ini mungkin mencakup daerah perkotaan atau subur dengan tingkat ekonomi yang tinggi. Kemudian pada kluster 3 menampilkan laju pertumbuhan penduduk (LPP) yang sedang sekitar 1.32%, dengan pengeluaran perkapita (PP) terendah di antara kluster lainnya, sekitar 21,651.28. Kepadatan penduduk (KP) relatif rendah, sekitar 1,520.556 orang per kilometer persegi, begitu juga dengan jumlah angkatan kerja (AK), yang juga relatif rendah, sekitar 752,103.5 orang. Kluster ini mungkin mencakup daerah pedesaan atau suburban dengan tingkat ekonomi yang lebih rendah.

4. Kesimpulan

Berdasarkan hasil analisis deskriptif, dapat disimpulkan bahwa Provinsi Jawa Barat menunjukkan pola laju pertumbuhan penduduk yang stabil, pengeluaran perkapita yang cenderung beragam, kepadatan penduduk yang umumnya moderat, dan jumlah Angkatan kerja yang signifikan, dengan variasi antara kabupaten-kabupaten yang ada di wilayah tersebut. Dalam analisis pengelompokkan dengan menggunakan 2 klaster, terdapat perbedaan yang cukup signifikan antara kedua kluster tersebut dalam hal karakteristik demografis dan ekonomi, yang memperlihatkan variasi dalam perkembangan dan kondisi sosio-ekonomi di berbagai kabupaten/kota di Jawa Barat. Secara umum, tingkat kesejahteraan masyarakat di Provinsi Jawa Barat pada tahun 2022 masih belum merata. Hal ini terlihat dari perbandingan antara klaster pertama dan klaster kedua dalam hal kepadatan penduduk, angkatan kerja, laju pertumbuhan penduduk, rata-rata pengeluaran perkapita, angka harapan hidup, dan rata-rata lama sekolah. Dengan membandingkan kedua hasil pengelompokkan tersebut, dapat dilihat bahwa penggunaan 3 klaster menghasilkan pengelompokkan yang lebih detail dibandingkan dengan 2 klaster. Dalam pengelompokkan dengan 3 klaster, terdapat perbedaan yang lebih jelas antara kelompok-kelompok kabupaten/kota. Oleh karena itu, hasil dari pengklasteran dengan metode k-means menyarankan bahwa kabupaten/kota yang tergolong dalam klaster ketiga membutuhkan lebih banyak bantuan dari pemerintah dibandingkan dengan kabupaten/kota yang tergolong dalam klaster pertama.

5. Daftar Pustaka

- [1] D. Sukmasari, "Konsep Kesejahteraan Masyarakat Dalam Perspektif Al-Qur'an," *Journal Of Qur'an and Hadis Studies*, vol. III, pp. 1-16, 2020.
- [2] D. Cahyadi, "Analisis Pengukuran Kesejahteraan Di Indonesia," *Jurnal Ilmiah*, 2017.
- [3] A. M. K. Mahmud and A. Nurfurkon, "Implementasi Peningkatan Kualitas Perencanaan Pembangunan Daerah Oleh Badan Perencanaan Pembangunan, Penelitian Dan Pengembangan Daerah," *Jurnal Pembangunan Pemberdayaan Pemerintahan*, vol. VIII, pp. 42-55, 2023.
- [4] B. P. S. P. D. Yogyakarta, 2022. [Online]. Available: <https://yogyakarta.bps.go.id/>.
- [5] A. A. Syahri, "Statistika Pendidikan," *Jurnal Pendidikan Matematika*, pp. 121-129, 2014.
- [6] L. M. Nasution, "Statistik Deskriptif," *Jurnal Hikmah*, pp. 49-55, 2017.
- [7] Kusnawi, "Pengantar Solusi Data Mining," 2007. [Online].

- [8] R. Febrian and A. Kesumawati, "Implementation Association Rules with Apriori Algorithm of Student Attendance Record in Islamic University of Indonesia," *International Journal of Applied Business and Information Systems*, vol. I, pp. 1-5, 2017.
- [9] M. B. I. Fuady and J. Nugraha, "Implementasi Metode K-Means Dan K-Medoids Untuk Mengelompokkan 82 Kota Di Indonesia Berdasarkan Indeks Harga Konsumen," p. 328.
- [10] B. S. Fadillah and J. Nugraha, "Komparasi Analisis K-Medoids Clustering dan Hierarchical clustering," p. 3, 2020.
- [11] S. N. Huda, "GIS-Based Provinces Clustering in Indonesia Based on the Production of Herbal Plants using K-Means," *American Scientific Publishers*, vol. 24, pp. 641-645, 2018.
- [12] Setiawati, "Analisis Pengaruh Kebijakan Deviden Terhadap Nilai Perusahaan Pada Perusahaan Farmasi Di BEI," *Jurnal Inovasi Penelitian*, pp. 1581-1590, 2021.
- [13] T. R. Tulili, K. B. Utomo and D. Lestari, "Pengelompokan Sekolah Dasar Di Kota Samarinda Berdasarkan Indikator Standar Pelayanan Minimal Pendidikan Menggunakan Metode K-Means," *JUST TI*, pp. 24-28, 2019.
- [14] Z. I. Alfianti, "Pengelompokan Tingkat Pendidikan Berdasarkan Jumlah Sekolah Di Provinsi Jawa Barat Menggunakan Algoritma K-Means," *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, pp. 33-39, 2022.
- [15] T. Khotimah, "Pengelompokan Surat Dalam Al-Qur'an Menggunakan Algoritma K-Means," *SIMETRIS*, pp. 83-88, 2014.
- [16] M. Sriningsih, . D. Hatidja and J. . D. Prang, "Penanganan Multikolinearitas Dengan Menggunakan Analisis Regresi Komponen Utama Pada Kasus Impor Beras Di Provinsi Sulut," *Jurnal Ilmiah Sains*, pp. 18-24, 2018.
- [17] M. Muhajir, *Modul Praktikum Statistika Multivariat Terapan*, Yogyakarta: Universitas Islam Indonesia, 2020.
- [18] A. R. BP, S. . A. Munandar, A. Fitriani, Y. Karlina and Y. ., "Pengertian Pendidikan, Ilmu Pendidikan Dan Unsur-Unsur Pendidikan," *journal.unismuh.ac.id*, pp. 1-8, 2022.
- [19] A. I. Permono, M. Alwi, B. Karunia, N. Michelle, A. . J. Pitoyo and . M. . A. F. Alfana, "Analisis Indeks Pembangunan Pendidikan Nusa Tenggara Barat 2019," 2020.
- [20] W. Lestari, "Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi," *SIMKOM*, pp. 35-48, 2019.
- [21] B. Santoso, *Data Mining : Teknik Pemanfaatan Data Untuk Keperluan Bisnis*, Yogyakarta: Graha Ilmu, 2007.
- [22] S. Y. Ghaisani, N. Hikmah, A. P. Hadi and E. Widodo, "Analisis Cluster Hirarki Untuk Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Demokrasi Indonesia Tahun 2016," pp. 1-11, 2019.
- [23] A. M. Tingginehe, J. O. Waani and C. E. Wuisang, "Perencanaan Pariwisata Hijau Di Distrik Roon Kabupaten Teluk Wondama, Papua Barat," *Jurnal Perencanaan Wilayah dan Kota*, vol. 6, pp. 511-520, 2019.