

Perbandingan Metode Klustering *K-Means* dan *DBSCAN* dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat

Siti Mutiah^{1*}, Yunia Hasnataeni², Anwar Fitrianto³, Erfiani⁴, L.M. Risman Dwi Jumansyah⁵

^{1,2,3,4,5}Statistika dan Sains Data, IPB University, Bogor, Jawa Barat, Indonesia

E-mail: ¹sitimutiah@apps.ipb.ac.id, ²yunia200698@gmail.com, ³anwarstat@gmail.com,

⁴erfiani@apps.ipb.ac.id, ⁵rismandwijumansyah@apps.ipb.ac.id

*Corresponding Author

ABSTRACT

This study aims to compare the effectiveness of two clustering methods, namely *K-Means* and *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)*, in grouping households based on socio-economic characteristics in West Java. The comparison of these two methods is important as each method has different strengths and limitations. *K-Means* excels in handling data with more homogeneous clusters, while *DBSCAN* is more flexible in managing outliers and irregular clusters that often appear in socio-economic data. The data used covers four categories: Household Facilities, Water Availability and Quality, Social and Economic Assistance, and Economic Welfare. The analysis results show disparities in access to facilities, clean water, and socio-economic assistance across various regions, with areas like Bandung and Garut outperforming Indramayu and Cirebon. The best method is determined by the highest silhouette score. The *K-Means* method produced more structured segmentation with a silhouette score of 0.69, indicating good performance in clustering data with more homogeneous characteristics. On the other hand, the *DBSCAN* method, which is more flexible in handling outliers, resulted in 7 clusters with 248 noise points and a lower silhouette score of 0.398, indicating a weaker cluster structure. The comparison of these two methods is relevant in the context of household clustering in West Java, where *K-Means* is more effective for areas with uniform facility access, while *DBSCAN* is better at capturing irregular variations and outliers. The explanation of the comparison between the two methods has been further elaborated to cover how variations in socio-economic access across different regions affect the effectiveness of each method, thus providing a deeper understanding of their respective strengths and limitations in handling the heterogeneity of socio-economic data in West Java.

Keywords: *DBSCAN*, *K-Means*, Clustering, Susenas

ABSTRAK

Penelitian ini bertujuan untuk membandingkan efektivitas dua metode klustering, yaitu *K-Means* dan *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)*, dalam mengelompokkan rumah tangga berdasarkan karakteristik sosial ekonomi di Jawa Barat. Perbandingan kedua metode ini penting karena masing-masing metode memiliki kelebihan dan keterbatasan yang berbeda, *K-Means* unggul dalam menangani data dengan kluster yang lebih seragam, sedangkan *DBSCAN* lebih fleksibel dalam mengelola *outlier* dan kluster tidak teratur yang sering muncul dalam data sosial ekonomi. Data yang digunakan meliputi empat kategori: Fasilitas Rumah Tangga, Ketersediaan dan Kualitas Air, Bantuan Sosial dan Ekonomi, serta Kesejahteraan Ekonomi. Hasil analisis menunjukkan ketimpangan dalam akses fasilitas, air bersih, dan bantuan sosial ekonomi di berbagai wilayah, di mana wilayah seperti Bandung dan Garut lebih unggul dibanding Indramayu dan Cirebon. Metode terbaik dilihat dari nilai silhouette tertinggi. Metode *K-Means* menghasilkan segmentasi yang lebih terstruktur dengan skor *silhouette* 0,69, menunjukkan performa yang baik dalam mengelompokkan data dengan karakteristik yang lebih seragam. Sebaliknya, metode *DBSCAN*, yang lebih fleksibel dalam menangani *outlier*, menghasilkan 7 kluster dengan 248 *noise points* dan skor *silhouette* yang lebih rendah yaitu 0,398, mengindikasikan struktur kluster yang kurang kuat. Perbandingan kedua metode ini relevan dalam konteks klustering rumah tangga di Jawa Barat, di mana *K-Means* lebih efektif untuk wilayah dengan akses fasilitas yang seragam, sedangkan *DBSCAN* lebih baik dalam menangkap variasi yang tidak beraturan dan *outlier*. Penjelasan perbandingan kedua metode ini telah diperinci lebih lanjut untuk mencakup bagaimana variasi akses sosial ekonomi di berbagai wilayah memengaruhi efektivitas masing-masing metode sehingga memberikan pemahaman yang lebih mendalam tentang keunggulan dan keterbatasan keduanya dalam menangani heterogenitas data sosial ekonomi di Jawa Barat.

Kata kunci: *DBSCAN*, *K-Means*, Klustering, Susenas

Dikirim: September 2024; Diterima: September 2024; Dipublikasikan: September 2024

Cara sitasi: Mutiah, S., Hasnateni, Y., Fitrianto, A., Erfiani & Jumansyah, L. M. R. D. (2024). Perbandingan Metode Klustering *K-Means* dan *DBSCAN* dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat. *Teorema: Teori dan Riset Matematika*, 09(02), 247-260. DOI: <http://dx.doi.org/10.25157/teorema.v9i2.16290>

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



PENDAHULUAN

Dalam era digital saat ini, jumlah data yang tersedia dari berbagai bidang, termasuk sosial dan ekonomi, terus meningkat dengan pesat. Peningkatan ini tidak hanya mempengaruhi cara kita mengumpulkan informasi tetapi juga menuntut metode analisis yang lebih unggul untuk mengelola dan menginterpretasikan data tersebut (Awad & Hamad, 2022). Salah satu teknik analisis data yang semakin mendapatkan perhatian adalah klustering. Klustering merupakan metode dalam data mining yang bertujuan untuk mengelompokkan objek-objek ke dalam beberapa kelompok atau kluster. Tujuannya adalah untuk mengidentifikasi dan mengelompokkan objek-objek yang memiliki kemiripan karakteristik ke dalam kluster yang sama, sementara objek yang berbeda ditempatkan dalam kluster yang berbeda (Suraya *et al.*, 2023; Setiawati *et al.*, 2024).

Di bidang sosial dan ekonomi, klustering memegang peranan yang sangat krusial. Teknik ini digunakan untuk mengungkap pola distribusi sumber daya, fasilitas, dan ketersediaan bantuan dalam masyarakat. Dengan analisis klustering, kita dapat memperoleh wawasan mendalam mengenai bagaimana sumber daya sosial dan ekonomi terdistribusi di berbagai kelompok masyarakat sehingga memungkinkan pemangku kebijakan untuk merancang intervensi yang lebih efektif dan terarah (Setiawati *et al.*, 2024; Nurhaliza & Mustakim., 2021). Di Jawa Barat, yang merupakan salah satu daerah dengan tantangan signifikan terkait ketidakmerataan akses terhadap fasilitas sosial dan ekonomi, klustering menjadi alat yang sangat penting. Teknik ini dapat digunakan untuk mengidentifikasi dan memahami kelompok rumah tangga berdasarkan akses mereka terhadap berbagai fasilitas sosial dan ekonomi, membantu dalam merancang strategi yang lebih tepat sasaran untuk mengatasi ketidakmerataan tersebut.

Dua metode klustering yang sering digunakan adalah *K-Means* dan *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)* (Cahapin *et al.*, 2023). Metode *K-Means* dikenal karena kesederhanaannya dan kemampuannya dalam mengelompokkan data berdasarkan kedekatan spasial relatif terhadap titik pusat kluster. *K-Means* secara efektif membagi data ke dalam sejumlah kluster yang telah ditentukan sebelumnya, yang menjadikannya cocok untuk dataset dengan struktur yang relatif homogen dan terdistribusi secara merata (Novianti *et al.*, 2017). Sebaliknya, *DBSCAN* yang merupakan metode berbasis kepadatan, menawarkan keuntungan tambahan dalam hal kemampuan untuk mengidentifikasi kluster dengan bentuk yang tidak teratur dan untuk menangani *outlier* dengan lebih efektif. *DBSCAN* tidak memerlukan jumlah kluster yang ditentukan sebelumnya, sehingga memberikan fleksibilitas yang lebih besar dalam menangani data yang memiliki distribusi yang lebih kompleks dan tidak teratur (Nurhaliza & Mustakim., 2021).

Pemilihan kedua metode dalam penelitian ini didasarkan pada kebutuhan untuk membandingkan efektivitas metode berbasis partisi, seperti *K-Means*, dengan metode berbasis kepadatan, seperti *DBSCAN*, dalam konteks klustering rumah tangga di Jawa Barat. *K-Means* unggul dalam memisahkan data ke dalam kluster yang seragam berdasarkan rata-rata kedekatan, sehingga memungkinkan pengelompokan yang lebih terstruktur. Namun, metode ini memiliki keterbatasan dalam menangani kluster yang tidak teratur, yang sering ditemukan pada data sosial ekonomi rumah tangga di Jawa Barat, terutama di daerah pedesaan dan perkotaan yang memiliki variasi fasilitas dan akses ekonomi yang signifikan (Setiawati *et al.*, 2024). Sebagai contoh, rumah tangga di daerah perkotaan seperti Kota Bandung dan Depok cenderung memiliki fasilitas lebih baik, sementara rumah tangga di daerah seperti Kabupaten seperti Sukabumi seringkali memiliki akses terbatas terhadap fasilitas dasar seperti air bersih dan bantuan sosial (Kesuma & Purwoto, 2022).

Di sisi lain, *DBSCAN* menawarkan fleksibilitas yang lebih besar dalam menangani data dengan bentuk kluster yang tidak beraturan dan dalam mengelola *outlier*. Dalam konteks rumah tangga di Jawa Barat, di mana terdapat variasi yang sangat luas dalam akses terhadap fasilitas dan bantuan sosial ekonomi, *DBSCAN* mampu mengidentifikasi kluster dengan bentuk yang lebih kompleks. Misalnya, rumah tangga di daerah dengan akses yang sangat terbatas, seperti Kabupaten Sukabumi, dapat diidentifikasi sebagai *outlier*

dalam klastering *K-Means*, tetapi dalam *DBSCAN*, rumah tangga ini dapat tetap dikelompokkan secara efektif berdasarkan kedekatan kepadatan data (Nurhaliza & Mustakim., 2021).

Secara keseluruhan, perbandingan antara kedua metode ini relevan untuk konteks Jawa Barat karena perbedaan karakteristik rumah tangga yang mencerminkan variasi dalam akses terhadap fasilitas sosial, air bersih, dan bantuan ekonomi. *K-Means* mungkin lebih cocok untuk daerah dengan akses yang lebih seragam, seperti perkotaan, sementara *DBSCAN* lebih efektif dalam menangkap heterogenitas di daerah pedesaan dan dalam mengatasi outlier yang tidak dapat dikelola dengan baik oleh *K-Means* (Setiawati *et al.*, 2024; Nurhaliza & Mustakim., 2021).

Tujuan dari penelitian ini adalah untuk membandingkan performa kedua metode klastering, yaitu *K-Means* dan *DBSCAN*, dalam mengelompokkan rumah tangga di Jawa Barat berdasarkan variabel ketersediaan fasilitas sosial ekonomi. Dengan melaksanakan analisis ini, diharapkan penelitian ini dapat memberikan wawasan yang lebih mendalam mengenai karakteristik kelompok rumah tangga di Jawa Barat. Hasil dari penelitian ini juga diharapkan dapat memberikan kontribusi yang signifikan terhadap perumusan kebijakan yang lebih tepat sasaran dan efektif dalam menangani ketidakmerataan akses terhadap fasilitas sosial dan ekonomi di wilayah tersebut.

METODE PENELITIAN

Data

Penelitian ini menggunakan data rumah tangga yang diperoleh dari Survei Sosial Ekonomi Nasional (Susenas) tahun 2023. Susenas, adalah survei yang dirancang oleh Badan Pusat Statistik (BPS) untuk mengumpulkan informasi mengenai kesejahteraan rumah tangga di Indonesia. Survei ini dilaksanakan setiap semester, yakni pada bulan Maret (Semester I) dan September (Semester II). Tujuan utama dari Susenas adalah untuk menyediakan data yang mencakup berbagai aspek kesejahteraan rumah tangga, termasuk pendidikan, kesehatan, dan daya beli. Dalam publikasi Statistik Kesejahteraan Rakyat Provinsi Jawa Barat 2023, data yang disajikan mencakup tabel-tabel yang menggambarkan statistik sosial ekonomi seperti kependudukan, pendidikan, kesehatan, dan perumahan. Data ini, yang diperoleh dari pelaksanaan Susenas Maret 2023 dengan melibatkan 26.040 sampel rumah tangga, disajikan pada tingkat provinsi dan kabupaten/kota, sehingga memungkinkan perbandingan antarwilayah. Data yang digunakan dalam penelitian ini terdiri dari beberapa variabel utama yang dikelompokkan ke dalam kategori berikut:

Tabel 1. Variabel yang digunakan dalam penelitian

Kategori	Variabel	Kode	Tipe Data
Fasilitas Rumah Tangga	Luas lantai rumah	(R1804)	Numerik
	Memiliki tabung gas	(R2001A)	Numerik
	Memiliki lemari es/kulkas	(R2001B)	Numerik
	Memiliki sepeda motor	(R2001H)	Numerik
Ketersediaan dan Kualitas Air	Sumber air utama untuk minum	(R1810A)	Numerik
	Waktu untuk mengambil air minum	(R1811B)	Numerik
	Kondisi fisik sumber air utama keruh	(R1813A)	Numerik
Bantuan Sosial dan Ekonomi	Menerima bantuan pangan di Februari 2023	(R2208A2)	Numerik
	Nilai bantuan yang diterima di Februari 2023	(R2208B2)	Numerik
	Menerima bantuan tunai rutin dari pemerintah daerah	(R2211A)	Numerik
Kesejahteraan Ekonomi	Menerima Kredit Usaha Rakyat (KUR)	(R1901A)	Numerik
	Memiliki usaha mikro atau kecil	(R2210A)	Numerik

Tahapan Penelitian

Penelitian ini dirancang untuk memberikan pemahaman mendalam mengenai pengelompokan data rumah tangga berdasarkan berbagai faktor sosial dan ekonomi di Jawa Barat. Tahapan penelitian ini terdiri

dari beberapa langkah kunci yang bertujuan untuk mengeksplorasi, menganalisis, dan mengevaluasi data secara sistematis sebagai berikut:

1. Eksplorasi Data

Pada tahap eksplorasi data, akan dilakukan analisis *univariat* dan *bivariat*. Dalam analisis *univariat*, masing-masing variabel akan dideskripsikan secara individu, seperti distribusi luas lantai rumah, kepemilikan fasilitas rumah tangga (misalnya tabung gas, kulkas, sepeda motor), sumber air utama, dan bantuan sosial yang diterima oleh rumah tangga. Selanjutnya, pada analisis *bivariat*, dilakukan pengujian hubungan antar variabel tersebut, misalnya dengan mengamati bagaimana kepemilikan fasilitas rumah tangga berkaitan dengan penerimaan bantuan sosial.

2. Analisis Kluster

Analisis kluster adalah metode dalam data mining yang bertujuan untuk mengelompokkan data (objek) ke dalam beberapa kategori atau kluster. Tujuan utamanya adalah mengumpulkan objek dengan pola yang mirip dalam satu kluster, sementara objek dengan pola yang berbeda ditempatkan dalam kluster yang berbeda (Möllensiep *et al.*, 2024). Dengan cara ini, klustering berupaya untuk mengoptimalkan kesamaan di dalam kluster yang sama dan memperjelas perbedaan antara kluster-kluster yang berbeda. Teknik ini berfokus pada pengelompokan data berdasarkan kesamaan strukturalnya, bukan pada prediksi data (Zhang *et al.*, 2024).

Analisis kluster mencakup dua pendekatan utama yaitu metode hierarkis dan non-hierarkis. Metode hierarkis dimulai dengan mengelompokkan data yang paling mirip satu sama lain dan kemudian melanjutkan ke objek yang memiliki kesamaan kedua terdekat, dan seterusnya. Proses ini menghasilkan struktur pohon yang disebut dendrogram, yang menunjukkan hierarki objek dari yang paling mirip hingga yang paling tidak mirip. Di sisi lain, metode non-hierarkis lebih fokus pada pengelompokan objek dalam jumlah tertentu tanpa membentuk struktur hierarkis. Dalam metode ini, jumlah kluster yang diinginkan (k), ditentukan oleh peneliti atau selama proses klustering itu sendiri. Metode non-hierarkis biasanya lebih efektif untuk digunakan pada sampel yang besar dibandingkan dengan metode hierarkis (Fahmiyah & Ningrum, 2023).

a. *K-Means*

K-Means merupakan salah satu metode pengelompokan non-hierarkis yang paling populer (Ridzki *et al.*, 2023). *K-Means* adalah algoritma pengelompokan yang sederhana dan sering digunakan. Algoritma ini bertujuan membagi data ke dalam beberapa kategori. Setelah memulai dengan pembagian data yang mungkin belum optimal, setiap titik data dipindahkan ke pusat kategori terdekat. Pusat kategori kemudian diperbarui dengan menghitung rata-rata posisi titik-titik tersebut, dan proses ini diulang hingga hasilnya stabil (Pranata *et al.*, 2022). Algoritma *K-Means* dalam metode klustering yang mengelompokkan titik data ke dalam grup yang serupa. Prosesnya melibatkan langkah-langkah berikut (Suraya *et al.*, 2023):

1. *Inisialisasi Centroid*: Mulailah dengan menentukan jumlah kluster (k) yang diinginkan dan pilih titik *centroid* secara acak sebagai titik awal klustering.
2. *Pengelompokan Data*: Tiap titik data akan dikelompokkan ke dalam grup yang memiliki pusat terdekat, dengan pusat grup diukur menggunakan jarak *euclidean*.

$$d(P, Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (1)$$

Dimana:

$d(P, Q)$: Jarak antara titik P ke titik Q

3. *Pembaruan Centroid*: Hitung *centroid* baru dengan mengambil rata-rata posisi titik-titik dalam setiap grup yang terbentuk. Proses ini dilakukan dengan menggunakan data yang sudah dikelompokkan.

$$u_j(t + 1) = \frac{1}{N_{sj}} \sum_{j \in s_j} x_j \quad (2)$$

Dimana:

$u_j(t + 1)$: Pusat kluster (centroid) yang baru pada iterasi ke - $t + 1$

N_{sj} : Jumlah data atau titik yang termasuk dalam kluster ke - j

4. Proses Iterasi: Ulangi langkah 2 dan 3 sampai tidak ada perubahan signifikan dalam penugasan kluster atau hingga mencapai jumlah iterasi yang ditentukan.
5. Evaluasi Konvergensi: Konvergensi dinilai dengan memantau perubahan pada pusat kluster atau dengan memeriksa apakah penugasan kluster tetap stabil dalam beberapa iterasi.
6. Penghentian: Algoritma berhenti ketika kriteria penghentian terpenuhi, seperti jumlah iterasi maksimum tercapai atau tidak ada perubahan signifikan dalam penugasan kluster.

b. *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)*

Algoritma *DBSCAN* adalah metode pengelompokan data spasial berdasarkan kepadatan yang diperkenalkan oleh Ester Martin pada tahun 1996 (Nurhaliza & Mustakim., 2021). Algoritma ini merupakan pengelompokan nonparametrik dalam pembelajaran tanpa pengawasan sehingga tidak memerlukan asumsi khusus dalam prosesnya. *DBSCAN* dapat membentuk kluster dengan bentuk bebas dan acak, dan efektif dalam mengatasi *noise* atau *outlier* yang mungkin ada dalam data. Algoritma ini bekerja dengan mengidentifikasi area yang memiliki kepadatan tinggi dan mengelompokkannya menjadi kluster. Dua parameter utama yang harus ditentukan adalah *radius* (ϵ) dan jumlah minimum objek (MinPts), yang digunakan untuk menentukan apakah suatu area dianggap padat atau tidak (Wulandari *et al.*, 2024). Berikut adalah algoritma *DBSCAN* (Rohalidyawati *et al.*, 2020):

1. Inisialisasi parameter input jumlah minimum objek (MinPts) dan *radius* (ϵ).
2. Pilih titik awal atau p secara acak.
3. Hitung jarak *radius* (ϵ) atau semua jarak titik yang dapat dijangkau dari p dengan menggunakan rumus jarak *euclidean*.

$$d(P, Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (3)$$

Dimana:

$d(P, Q)$: Jarak antara titik P ke titik Q

4. Jika jumlah titik yang memenuhi kriteria *radius* (ϵ) lebih dari MinPts, maka titik p dianggap sebagai titik inti dan membentuk sebuah kluster.
5. Ulangi langkah 3 dan 4 untuk semua titik. Jika p adalah titik border dan tidak ada titik lain yang dapat dijangkau dari p , maka lanjutkan proses ke titik berikutnya.

3. Visualisasi Hasil Klustering

Visualisasi hasil klustering melibatkan beberapa pendekatan. Untuk visualisasi klustering *K-Means*, digunakan grafik *density plot* yang menampilkan kluster-kluster yang terbentuk dengan warna berbeda, serta memperlihatkan posisi *centroid* dari masing-masing kluster. Visualisasi ini membantu memahami distribusi rumah tangga dalam ruang multidimensi. Selanjutnya, pada visualisasi *DBSCAN*, kluster dan *outlier* yang terdeteksi ditampilkan menggunakan grafik *density plot* yang menunjukkan kepadatan data, dengan warna berbeda untuk membedakan kluster serta menyoroti *outlier* (Farahnakian *et al.*, 2023). Selain itu, peta tematik

digunakan untuk memvisualisasikan distribusi kluster di wilayah Jawa Barat, memberikan gambaran geografis mengenai sebaran rumah tangga dengan fasilitas sosial ekonomi yang berbeda.

4. Evaluasi Model

Evaluasi model dilakukan menggunakan *Indeks Silhouette (SI)*, salah satu ukuran validasi berbasis kriteria internal. *indeks silhouette* mengevaluasi kualitas pengelompokan dengan membandingkan jarak rata-rata antara objek di dalam kluster yang sama dan objek di kluster yang berbeda. Perhitungan nilai *indeks silhouette* menggunakan persamaan yang mempertimbangkan jarak rata-rata antar objek dalam kluster yang sama serta jarak rata-rata objek ke kluster terdekat yang berbeda (Adha *et al.*, 2021).

$$SI = \frac{1}{n} \sum_{i=1}^n \left(\frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \right) \quad (4)$$

Dimana:

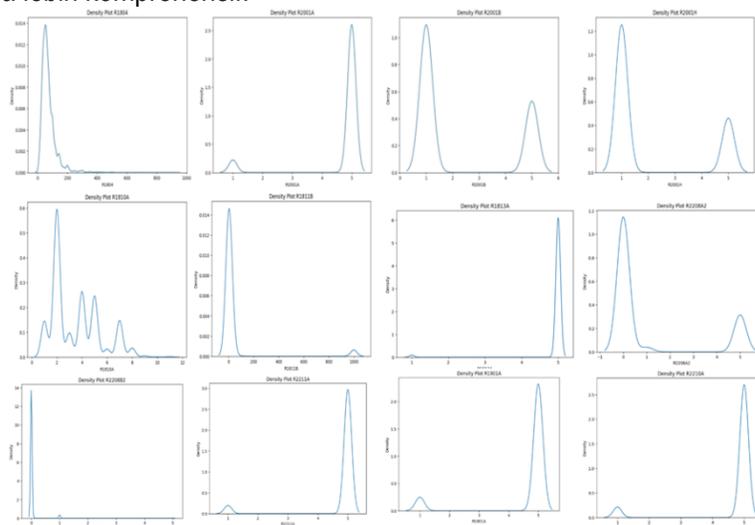
$a(i)$: Jarak rata-rata objek ke semua objek lain dalam kluster yang sama

$b(i)$: Jarak rata-rata objek ke objek-objek di kluster terdekat yang berbeda

HASIL DAN PEMBAHASAN

1. Eksplorasi Data

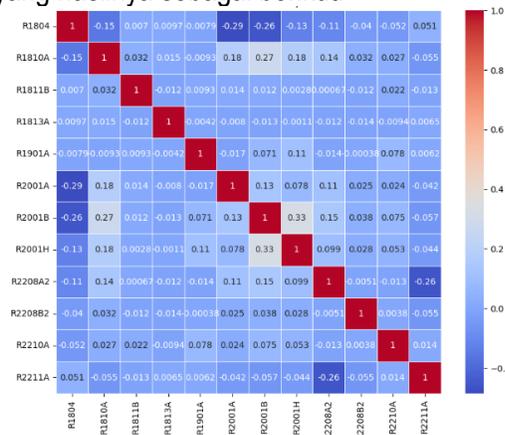
Penelitian ini menggunakan empat kategori data, yaitu: Fasilitas Rumah Tangga yang mencakup empat variabel (luas lantai rumah, kepemilikan tabung gas, kulkas, dan sepeda motor); ketersediaan dan kualitas air yang meliputi tiga variabel (sumber air utama untuk minum, waktu yang diperlukan untuk mengambil air, serta kondisi fisik sumber air yang keruh); bantuan sosial dan ekonomi dengan tiga variabel (penerimaan bantuan pangan, nilai bantuan yang diterima, dan penerimaan bantuan tunai rutin); serta kesejahteraan ekonomi yang terdiri dari dua variabel (penerimaan kredit usaha rakyat dan kepemilikan usaha mikro). Selain keempat kategori tersebut, data Provinsi dan kabupaten juga digunakan untuk analisis sebaran berdasarkan wilayah di Jawa Barat. Kami juga menggabungkan data Susenas dengan data spasial wilayah Jawa Barat, yang memungkinkan kami untuk menganalisis hubungan antara faktor sosial-ekonomi dan distribusi geografis secara lebih mendalam, sehingga dapat mengkaji keterkaitan kondisi sosial-ekonomi dengan lokasi secara lebih komprehensif.



Gambar 1. Eksplorasi setiap variabel

Eksplorasi setiap variabel menunjukkan variasi dalam Fasilitas Rumah Tangga, yaitu adanya kesenjangan dalam kepemilikan aset penting, seperti luas lantai rumah, kulkas, dan sepeda motor, yang mencerminkan tingkat kenyamanan hidup rumah tangga. Selain itu, eksplorasi pada Ketersediaan dan Kualitas Air menunjukkan bahwa meskipun sebagian rumah tangga memiliki akses terhadap sumber air bersih, masih ada rumah tangga yang membutuhkan waktu lebih lama untuk mendapatkan air, dan sebagian lainnya menghadapi masalah kualitas air yang keruh. Pada variabel Bantuan Sosial dan Ekonomi, hasilnya menunjukkan bahwa sebagian besar rumah tangga menerima bentuk bantuan pangan maupun tunai, meskipun nilai bantuan yang diterima bervariasi antar rumah tangga. Di sisi lain, dalam Kesejahteraan Ekonomi, eksplorasi menunjukkan bahwa akses terhadap kredit usaha rakyat dan kepemilikan usaha mikro belum merata, mengindikasikan adanya peluang untuk meningkatkan akses terhadap dukungan finansial guna mendorong kemandirian ekonomi.

Sebuah penelitian oleh Puspitasari *et al.*, (2024) mengenai pengelompokan daerah di Provinsi Jawa Barat berdasarkan indikator kesejahteraan masyarakat menggunakan analisis klustering, menunjukkan adanya kesenjangan dalam berbagai indikator kesejahteraan seperti kemiskinan, pendidikan, dan kesehatan yang mengindikasikan perlunya perbaikan di beberapa daerah. Penelitian ini mendukung temuan kami mengenai kesenjangan sosial-ekonomi yang ada, serta pentingnya intervensi yang terfokus untuk meningkatkan kualitas hidup masyarakat. Selain melakukan eksplorasi masing-masing variabel, kami juga membuat korelasi antar variabel yang hasilnya sebagai berikut:

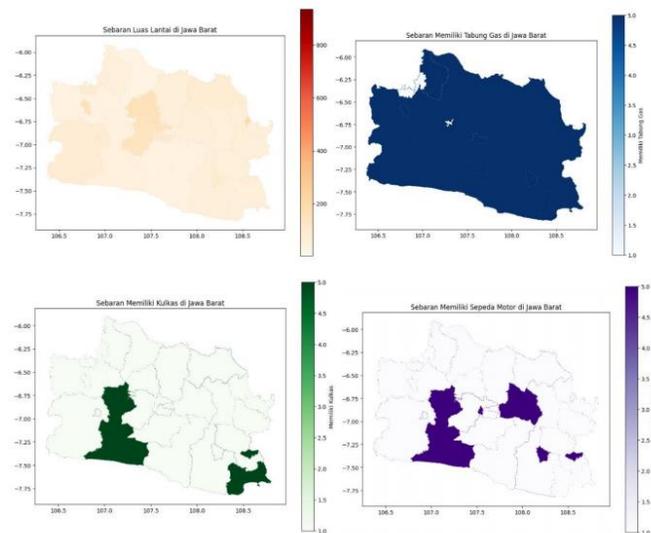


Gambar 2. Korelasi antar variabel

Hubungan antarvariabel Fasilitas Rumah Tangga menunjukkan korelasi negatif, seperti luas lantai rumah yang lebih besar tidak selalu diikuti kepemilikan tabung gas atau kulkas. Pada Ketersediaan dan Kualitas Air, korelasi rendah antarvariabel menunjukkan variasi akses dan kualitas air yang signifikan tanpa pola yang jelas. Sementara itu, dalam Bantuan Sosial dan Ekonomi, terdapat korelasi positif antara penerimaan bantuan pangan dan jumlah bantuan yang diterima, menunjukkan bahwa rumah tangga yang menerima bantuan cenderung memperoleh jumlah lebih besar. Namun, pada Kesejahteraan Ekonomi, korelasi antara penerimaan Kredit Usaha Rakyat (KUR) dan kepemilikan usaha mikro atau kecil masih lemah, menandakan bahwa penerimaan KUR tidak selalu berdampak langsung pada pengembangan usaha rumah tangga. Penelitian oleh Purnama (2013) menunjukkan bahwa meskipun kredit bank secara umum berkontribusi pada pertumbuhan ekonomi, dampaknya terhadap kesejahteraan rumah tangga sering kurang signifikan, sehingga diperlukan pendekatan kebijakan ekonomi yang lebih komprehensif, mencakup investasi, infrastruktur, dan dukungan sosial. Untuk memperjelas sebaran data, kami menggunakan peta Jawa Barat yang menggabungkan data spasial dengan data rumah tangga guna memvisualisasikan distribusi fasilitas dan bantuan di setiap daerah.



Gambar 3. Peta Jawa Barat

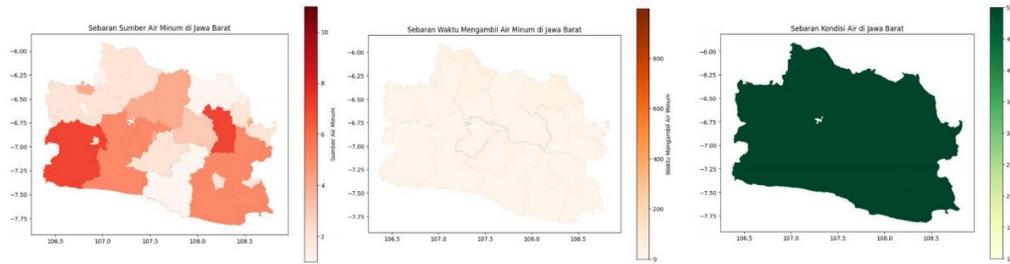


Gambar 4. Sebaran Fasilitas Rumah Tangga Jawa Barat

Berdasarkan peta sebaran fasilitas rumah tangga di Jawa Barat, terdapat perbedaan distribusi fasilitas di berbagai wilayah. Untuk luas lantai rumah, wilayah yang memiliki rumah dengan luas lantai lebih besar cenderung berada di bagian Barat Daya Jawa Barat, terutama di sekitar Kabupaten Bandung, Kabupaten Bandung Barat, dan Kabupaten Garut. Sebaliknya, daerah seperti Indramayu, Majalengka, dan beberapa bagian Bekasi memiliki luas lantai rumah yang lebih kecil. Hal ini sejalan dengan penelitian Kesuma & Purwoto (2022) yang menyebutkan bahwa kepemilikan luas lantai terluas ada di daerah selatan Jawa Barat. Pada kepemilikan tabung gas, sebarannya terlihat cukup merata di seluruh provinsi. Wilayah seperti Bogor, Depok, Bekasi, Indramayu, Cirebon, serta beberapa bagian Sukabumi dan Cianjur menunjukkan tingkat kepemilikan tabung gas yang tinggi.

Kepemilikan kulkas lebih dominan di wilayah Bandung dan Cimahi, yang memiliki warna hijau gelap pada peta, serta di Kabupaten Garut dan sebagian kecil wilayah Sukabumi. Namun, wilayah bagian utara seperti Indramayu, Cirebon, serta beberapa daerah di sekitar Subang dan Karawang menunjukkan tingkat kepemilikan kulkas yang lebih rendah. Sementara itu, kepemilikan sepeda motor tinggi di bagian tengah dan selatan Jawa Barat, khususnya di Kabupaten Bandung, Kabupaten Bandung Barat, Kabupaten Garut, dan sebagian Tasikmalaya. Sebaliknya, daerah utara seperti Indramayu, Cirebon, serta beberapa wilayah di Bekasi dan Depok memiliki tingkat kepemilikan sepeda motor yang lebih rendah. Secara keseluruhan,

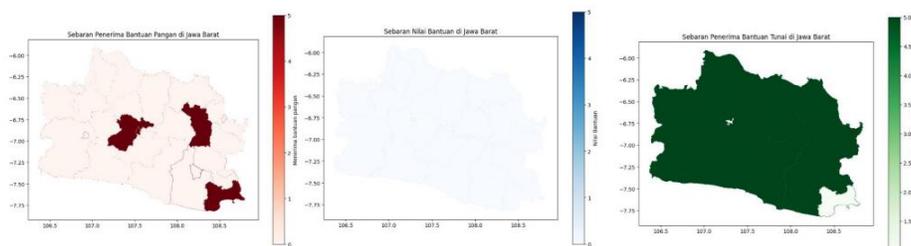
wilayah Barat Daya Jawa Barat seperti Bandung, Cimahi, dan Garut menunjukkan kecenderungan memiliki fasilitas rumah tangga yang lebih baik, sementara wilayah utara seperti Indramayu dan Cirebon cenderung memiliki akses yang lebih terbatas. Hal ini mencerminkan adanya kesenjangan sosial-ekonomi antar wilayah di provinsi tersebut.



Gambar 5. Sebaran Ketersediaan dan Kualitas Air Tangga Jawa Barat

Berdasarkan peta sebaran ketersediaan dan kualitas air di Jawa Barat yang mencakup tiga variabel, yaitu sumber air utama untuk minum, waktu yang diperlukan untuk mengambil air, dan kondisi fisik sumber air yang keruh, terdapat perbedaan kondisi antar wilayah. Pada variabel sumber air minum utama, wilayah yang lebih banyak menggunakan sumber air seperti sumur atau air tanah berada di selatan Bandung dan Kabupaten Garut, dengan warna merah tua yang menandakan ketergantungan tinggi terhadap sumber air tersebut. Di bagian utara dan timur Jawa Barat, seperti Indramayu, Cirebon, dan Subang, penggunaan sumber air utama lebih rendah. Sementara itu, variabel waktu yang diperlukan untuk mengambil air menunjukkan bahwa sebagian besar wilayah Jawa Barat memiliki akses yang mudah terhadap air minum, terlihat dari warna terang pada peta, yang menandakan waktu pengambilan air yang singkat. Tidak ada daerah yang memerlukan waktu lama untuk mengambil air.

Selanjutnya, pada variabel kondisi fisik sumber air, wilayah dengan kondisi air yang lebih bersih, ditunjukkan oleh warna hijau tua, tersebar di bagian utara dan selatan Jawa Barat, seperti di Bogor, Sukabumi, dan sebagian Bandung. Di sisi lain, wilayah bagian tengah Jawa Barat menunjukkan beberapa daerah dengan kondisi air yang lebih keruh, meskipun tidak signifikan. Secara keseluruhan, wilayah seperti Bandung, Garut, dan Sukabumi memiliki akses air minum yang lebih baik, baik dari segi ketersediaan, waktu akses, maupun kualitas fisik air. Hal ini mencerminkan bahwa sebagian besar wilayah Jawa Barat memiliki ketersediaan air minum yang baik dan kualitas air yang memadai, dengan beberapa daerah yang perlu perhatian lebih terhadap kebersihan airnya. Hal ini mencerminkan bahwa sebagian besar wilayah Jawa Barat memiliki ketersediaan air minum yang baik dan kualitas air yang memadai, dengan beberapa daerah yang perlu perhatian lebih terhadap kebersihan airnya, dan hal ini telah ditunjukkan oleh penelitian (Kesuma & Purwoto, 2022) yang menunjukkan capaian tertinggi air bersih di Jawa Barat mencapai 100 persen, jauh di atas wilayah capaian terendah sebesar 79,98 persen.



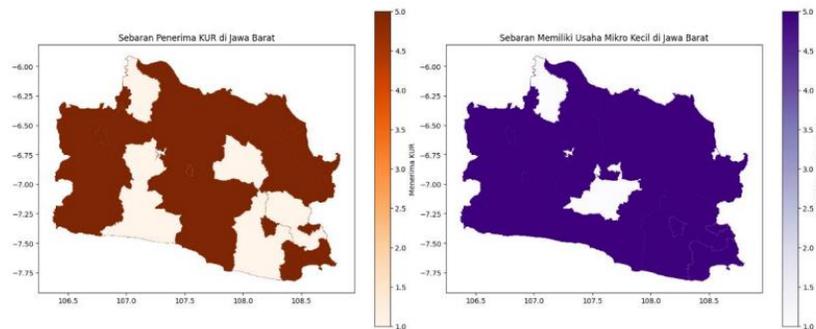
Gambar 6. Sebaran Bantuan Sosial Ekonomi Air Jawa Barat

Berdasarkan peta sebaran bantuan sosial dan ekonomi di Jawa Barat yang mencakup tiga variabel, yaitu penerimaan bantuan pangan, nilai bantuan yang diterima, dan penerimaan bantuan tunai rutin, terdapat beberapa perbedaan distribusi di berbagai wilayah. Pada variabel penerimaan bantuan pangan, terlihat

bahwa daerah yang paling banyak menerima bantuan pangan berada di wilayah tengah seperti Kabupaten Bandung dan sekitarnya serta bagian selatan di Kabupaten Garut. Wilayah-wilayah ini ditandai dengan warna merah tua, yang menunjukkan bahwa mereka merupakan penerima bantuan pangan dengan jumlah yang cukup besar dibandingkan dengan wilayah lain, yang cenderung berwarna lebih terang.

Untuk variabel nilai bantuan yang diterima, sebarannya relatif lebih seragam di seluruh wilayah Jawa Barat, dengan peta yang didominasi oleh warna yang sangat terang. Hal ini menunjukkan bahwa nilai bantuan yang diterima relatif merata di seluruh wilayah, tanpa ada wilayah yang menonjol dalam hal jumlah nilai bantuan yang diterima. Pada variabel penerimaan bantuan tunai rutin, wilayah dengan penerima terbanyak terlihat di bagian utara dan selatan Jawa Barat. Wilayah seperti Bogor, Sukabumi, dan sebagian Bandung menunjukkan penerimaan bantuan tunai rutin yang lebih tinggi, ditandai dengan warna hijau tua. Hal ini menunjukkan bahwa wilayah-wilayah tersebut memiliki lebih banyak penerima bantuan tunai rutin dibandingkan wilayah lain yang berwarna lebih terang.

Secara keseluruhan, peta ini memberikan gambaran bahwa wilayah-wilayah di sekitar Bandung dan Garut cenderung lebih banyak menerima bantuan pangan, sementara penerimaan bantuan tunai tersebar lebih luas di wilayah utara dan selatan Jawa Barat. Nilai bantuan yang diterima terlihat merata di seluruh provinsi, mencerminkan distribusi yang cukup seimbang dalam hal jumlah bantuan yang diterima oleh penduduk di berbagai wilayah. Bantuan pemerintah mampu memberikan pengaruh positif pada keadaan ekonomi masyarakat. Hasil ini sejalan dengan temuan Wahyuni *et al.*, (2023) yang menunjukkan pentingnya bantuan tunai dalam mendukung ketahanan ekonomi rumah tangga.



Gambar 7. Sebaran Kesejahteraan Ekonomi Tangga Jawa Barat

Pada peta diatas terlihat distribusi dua variabel penting di Jawa Barat, yaitu penerimaan Kredit Usaha Rakyat (KUR) dan kepemilikan usaha mikro. Wilayah yang terwarnai gelap pada peta penerimaan KUR, seperti Bandung, Bekasi, dan Depok, menunjukkan tingkat penerimaan KUR yang tinggi, yang mungkin mengindikasikan adanya akses yang lebih baik ke modal serta aktivitas ekonomi yang lebih dinamis di daerah tersebut. Sementara itu, daerah dengan warna lebih terang seperti Pangandaran, Sukabumi, dan Ciamis, mungkin mengalami kendala dalam mengakses fasilitas kredit.

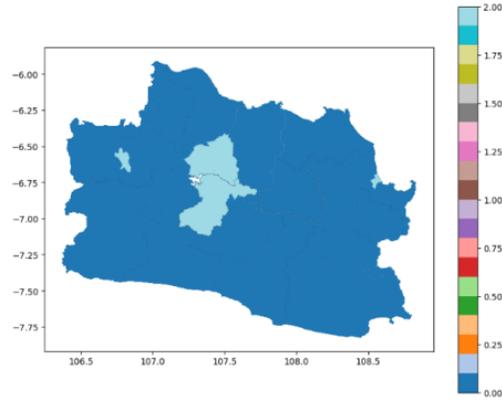
Untuk peta kepemilikan usaha mikro, daerah dengan warna gelap seperti Cirebon, Tasikmalaya, dan Kuningan, memiliki tingkat kepemilikan usaha mikro yang tinggi, menandakan adanya semangat kewirausahaan yang kuat atau dukungan yang baik dari pemerintah lokal untuk pengusaha kecil. Namun, daerah dengan warna lebih terang seperti Purwakarta dan Karawang, mungkin kurang dalam hal jumlah usaha mikro, mungkin karena fokus lebih kepada industri besar atau pekerjaan formal.

Analisis ini menunjukkan bahwa ada variasi yang signifikan dalam akses ke modal dan kegiatan wirausaha di seluruh provinsi, dengan beberapa daerah menunjukkan sinergi antara kredit usaha dan aktivitas ekonomi mikro sementara daerah lain mungkin memiliki potensi yang belum dimanfaatkan atau menghadapi hambatan dalam mengembangkan usaha kecil. Peta administratif membantu dalam mengaitkan

data ini dengan lokasi geografis spesifik, memberikan wawasan yang lebih dalam tentang dinamika ekonomi regional di Jawa Barat.

2. Analisis Kluster

Berikut adalah hasil klustering menggunakan metode *K-Means*:

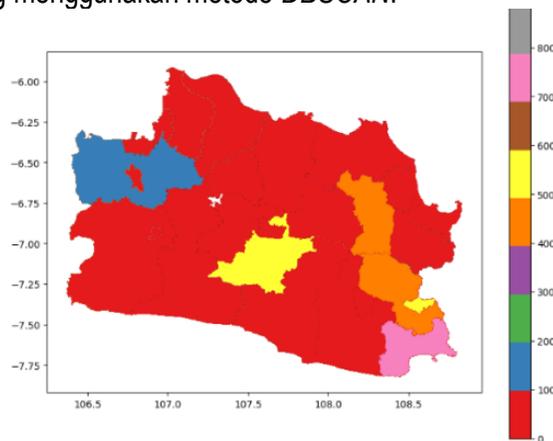


Gambar 8. Metode K-Means Klustering

Klustering K-means mengelompokkan data berdasarkan kesamaan fitur-fitur yang ada, dimana setiap warna pada peta mewakili satu kluster dengan karakteristik khusus yang membedakannya dari kluster lain. Daerah yang berwarna biru tua mungkin menandakan area dengan karakteristik unik yang berbeda signifikan dari daerah lain, sedangkan wilayah dengan warna biru muda dan hijau cenderung menunjukkan homogenitas dalam beberapa aspek. Daerah yang berwarna lebih cerah, seperti kuning hingga merah, mungkin mengindikasikan transisi atau perbedaan lain dalam karakteristik yang diamati.

Analisis ini sangat berguna dalam perencanaan pembangunan, mengalokasikan sumber daya dengan lebih efektif, dan untuk melakukan investigasi lebih lanjut tentang penyebab pembentukan kluster. Dengan memahami karakteristik dari tiap kluster yang terbentuk, pemangku kepentingan dapat lebih tepat dalam membuat keputusan terkait implementasi kebijakan atau intervensi yang ditujukan untuk memperbaiki kondisi di daerah tertentu berdasarkan data dan bukti yang diperoleh melalui metode klustering ini.

Berikut adalah hasil klustering menggunakan metode *DBSCAN*:



Gambar 9. Metode DBSCAN Klustering

Dari gambar hasil klustering DBSCAN yang Anda bagikan, terlihat bahwa wilayah peta telah dibagi menjadi beberapa kluster dengan warna yang berbeda, menandakan wilayah-wilayah yang memiliki karakteristik atau pola serupa berdasarkan metode DBSCAN. Area yang berwarna abu-abu mewakili noise

points, yang tidak terklasifikasi ke dalam kluster manapun. Dari visualisasi ini, dapat diinterpretasikan bahwa sebagian besar wilayah dipenuhi oleh satu kluster dominan (merah), sementara beberapa kluster lain (seperti biru, kuning, dan pink) memiliki area yang lebih kecil. Hal ini menunjukkan distribusi geografis yang bervariasi dalam data yang dianalisis.

3. Evaluasi Model

Hasil clustering K-Means menunjukkan nilai Inertia (SSE) sebesar 33.262.072, yang menunjukkan jumlah total kesalahan kuadrat antara setiap titik data dan pusat kluster terdekatnya. Nilai inertia yang lebih rendah umumnya menunjukkan bahwa titik-titik data dalam kluster lebih dekat ke pusatnya, namun nilai ini tidak dapat berdiri sendiri dan perlu dibandingkan dengan inertia dari model lain untuk evaluasi lebih lanjut. Sementara itu, Silhouette Score sebesar 0,69 menunjukkan bahwa clustering yang dihasilkan cukup baik, dengan nilai mendekati 1 yang mengindikasikan bahwa data secara umum dikelompokkan dengan benar ke dalam kluster yang tepat, dan jaraknya jauh dari kluster lainnya. Secara keseluruhan, hasil ini mengindikasikan bahwa K-Means memberikan segmentasi yang cukup kuat pada data yang dianalisis.

Dari hasil DBSCAN, terbentuk 7 kluster dengan 248 noise points, yaitu data yang tidak dimasukkan ke dalam kluster manapun. Jumlah noise points yang cukup besar ini menunjukkan bahwa data memiliki beberapa titik yang tidak sesuai dengan pola yang dikenali oleh algoritma. Skor Silhouette sebesar 0.398 mengindikasikan bahwa struktur clustering tidak terlalu kuat, karena nilainya mendekati 0. Ini menandakan bahwa beberapa titik mungkin berada di tepi kluster atau tidak jelas terasosiasi dengan kluster tertentu, sehingga perlu dipertimbangkan untuk mengevaluasi parameter atau menggunakan metode clustering lain untuk hasil yang lebih optimal.

Hasil ini sejalan dengan penelitian Budiman *et al.*, (2016) yang membandingkan metode K-Means dan DBSCAN. Penelitian tersebut menemukan bahwa metode K-Means lebih baik dalam klasifikasi dibandingkan DBSCAN, dengan nilai Silhouette index K-Means mencapai 0,463, lebih tinggi daripada nilai DBSCAN yang hanya 0,281. Selain itu, penelitian terkait pengelompokan data penyebaran Covid-19 menunjukkan bahwa K-Means juga lebih optimal, memiliki nilai DBI lebih baik dibandingkan DBSCAN (Biantara *et al.*, 2023).

KESIMPULAN

Penelitian ini mengidentifikasi perbedaan signifikan dalam distribusi fasilitas rumah tangga, akses air, bantuan sosial-ekonomi, dan kesejahteraan ekonomi di Jawa Barat. Eksplorasi data menunjukkan adanya ketimpangan dalam akses terhadap fasilitas rumah tangga seperti luas lantai, kepemilikan tabung gas, kulkas, dan sepeda motor. Beberapa wilayah seperti Kabupaten Bandung dan Kabupaten Garut memiliki akses lebih baik, sementara wilayah utara seperti Indramayu dan Cirebon menunjukkan akses yang lebih terbatas. Dalam akses air bersih, sebagian besar wilayah memiliki ketersediaan yang baik, namun beberapa wilayah masih menghadapi masalah waktu pengambilan air dan kualitas air yang keruh. Penerimaan bantuan sosial ekonomi juga menunjukkan variasi. Wilayah seperti Bandung dan Garut menerima lebih banyak bantuan pangan, sementara penerimaan bantuan tunai relatif merata di seluruh wilayah. Akses terhadap Kredit Usaha Rakyat (KUR) dan kepemilikan usaha mikro juga bervariasi di berbagai wilayah, dengan wilayah perkotaan seperti Bandung, Bekasi, dan Depok memiliki akses lebih baik dibandingkan daerah pedesaan seperti Sukabumi dan Pangandaran. Hal ini mencerminkan ketidakmerataan dalam kemandirian ekonomi di Jawa Barat.

Klustering menggunakan metode *K-Means* dan *DBSCAN* memberikan wawasan yang mendalam mengenai pembentukan kelompok berdasarkan karakteristik sosial ekonomi rumah tangga. *K-Means* menunjukkan segmentasi yang kuat dengan skor *silhouette* sebesar 0,69, menandakan klustering yang cukup baik. Sebaliknya, metode *DBSCAN* menghasilkan 7 kluster dengan 248 *noise points*, dan skor

silhouette yang lebih rendah yaitu 0,398, menunjukkan bahwa struktur kluster kurang kuat. Temuan ini mengindikasikan bahwa *K-Means* lebih efektif dalam mengelompokkan data social ekonomi di wilayah dengan karakteristik yang lebih seragam.

REKOMENDASI

Untuk meningkatkan hasil analisis, disarankan melakukan optimasi pada metode *K-Means* dengan mencoba variasi jumlah kluster untuk menemukan hasil yang lebih optimal. Selain itu, perlu dilakukan tuning pada parameter *radius* (ϵ) dan jumlah minimum objek (MinPts) pada metode DBSCAN guna mengurangi noise dan meningkatkan kualitas clustering. Penggunaan metrik validasi tambahan seperti Davies-Bouldin Index juga direkomendasikan untuk memberikan evaluasi yang lebih mendalam terhadap kualitas kluster yang dihasilkan. Terakhir, untuk memaksimalkan pemahaman terhadap hubungan antar variabel, analisis korelasi dapat diperluas dengan uji statistik tambahan.

UCAPAN TERIMAKASIH

Penelitian yang kami lakukan tidak dapat terlaksana apabila tanpa dukungan dari Departemen Statistika dan Sains Data IPB University. Maka dari itu, kami mengucapkan terima kasih kepada Departemen Statistika dan Sains Data IPB yang telah memberikan dukungan penuh dan fasilitas yang sangat membantu dalam menunjang penelitian ini.

DAFTAR PUSTAKA

- Adha, R., Nurhaliza, N., & Soleha, U. (2021). Perbandingan algoritma DBSCAN dan K-Means clustering untuk pengelompokan kasus Covid-19 di dunia. *Jurnal Sains, Teknologi Dan Industri*, 18(2), 206–211.
- Awad, F. H., & Hamad, M. M. (2022). Improved k-Means clustering algorithm for big data based on distributed smartphone neural engine processor. *Electronics*, 11(6), 1-20.
- Biantara, B., Rohana, T., & Ratna Juwita, A. (2023). Perbandingan Algoritma K-Means dan DBSCAN untuk Pengelompokan Data Penyebaran Covid-19 Seluruh Kecamatan di Provinsi Jawa Barat. *Scientific Student Journal for Information, Technology and Science*, IV(1), 88–94.
- Cahapin, E. L., Malabag, B. A., Santiago, C. S., Reyes, J. L., Legaspi, G. S., & Adrales, K. L. (2023). Clustering of students admission data using k-means, hierarchical, and DBSCAN algorithms. *Bulletin of Electrical Engineering and Informatics*, 12(6), 3647–3656. <https://doi.org/10.11591/eei.v12i6.4849>
- Fahmiyah, I., & Ningrum, R. A. (2023). Human development clustering in Indonesia: Using K-Means method and based on Human Development Index categories. *Journal of Advanced Technology and Multidiscipline*, 2(1), 27–33.
- Farahnakian, F., Nicolas, F., Farahnakian, F., Nevalainen, P., Sheikh, J., Heikkonen, J., & Raduly-Baka, C. (2023). A comprehensive study of clustering-based techniques for detecting abnormal vessel behavior. *Remote Sensing*, 15(6), 1-34.
- Kesuma, R., & Purwoto, A. (2022). Pengelompokan Kabupaten/Kota Berdasarkan Indikator Rumah Layak Huni di Provinsi Jawa Barat Tahun 2020. *Seminar Nasional Official Statistics*, 2022(1), 995–1004. <https://doi.org/10.34123/semnasoffstat.v2022i1.1303>
- Möllensiep, D., Schäfer, J., Pasch, F., & Kuhlenkötter, B. (2024). Cluster analysis for systematic database extension to improve machine learning performance in double-sided incremental sheet forming. *International Journal of Advanced Manufacturing Technology*, 133(9–10), 4301–4315.
- Novianti, P., Setyorini, D., & Rafflesia, U. (2017). K-means cluster analysis in earthquake epicenter clustering. *International Journal of Advances in Intelligent Informatics*, 3(2), 87–89
- Nurhaliza, N., & Mustakim. (2021). Clustering of data Covid-19 cases in the world using DBSCAN algorithms. *IJRSE: Indonesian Journal of Informatic Research and Software Engineering*, 1(1), 1–8.
- Pranata, K. S., Gunawan, A. A. S., & Gaol, F. L. (2022). Development clustering system IDX company with

- K-means algorithm and DBSCAN based on fundamental indicator and ESG. *Procedia Computer Science*, 216, 319–327.
- Purnama, E. S. (2013). *Pertumbuhan Ekonomi (Studi Kasus Pada 12 Provinsi Di Indonesia)*. 1–11.
- Puspitasari, P. A., Faidah, D. Y., & Hendrawati, T. (2024). *Grouping Regencies / Cities In West Java Province Based On People ' S Welfare Indicators Using Biplot And Clustering*. 18(3), 1839–1852.
- Ridzki, M. M., Hadijah, I., Mukidin, M., Azzahra, A., & and Nurjanah, A. (2023). K-Means algorithm method for clustering best-selling product data at XYZ grocery stores. *International Journal of Social Service and Research*, 3(12), 3354–3367.
- Rohalidyawati, W., Rahmawati, R., & Mustafid, M. (2020). Segmentasi pelanggan e-money dengan menggunakan algoritma DBSCAN (Density Based Spatial Clustering Applications With Noise) di Provinsi DKI Jakarta. *Jurnal Gaussian*, 9(2), 162–169.
- Setiawati, E., Fernanda, U. D., & Agesti, S. (2024). Implementation of K-Means , K-Medoid and DBSCAN Algorithms In Obesity Data Clustering. *IJATIS: Indonesian Journal of Applied Technology and Innovation Science*, 23–29. <https://journal.irpi.or.id/index.php/ijatis>
- Sisca Agustin Diani Budiman, Diah Safitri, & Dwi Ispriyanti. (2016). Perbandingan Metode K-Means dan Metode DBSCAN Pada Pengelompokan Rumah Kost Mahasiswa di Kelurahan Tembalang Semarang. *Jurnal Gaussian*, 5(4), 757–762.
- Suraya, S., Sholeh, M., & and Lestari, U. (2023). Evaluation of data clustering accuracy using K-Means algorithm. *International Journal of Multidisciplinary Approach Research and Science*, 2(1), 385–396.
- Wahyuni, W., Dwiarto, R., Suwarno, R. S., & Giyanto, B. (2023). Evaluasi Kebijakan Perlindungan Sosial dan Pengentasan Kemiskinan Melalui Program Keluarga Harapan (PKH). *Jurnal Pembangunan Dan Administrasi Publik*, 5(2), 11–22.
- Wulandari, V., Syarif, Y., Alfian, Z., Althof, M. A., & Mufidah, M. (2024). Comparison of density-based spatial clustering of applications with noise (DBSCAN), K-Means and X-Means algorithms on shopping trends data. *IJATIS: Indonesian Journal of Applied Technology and Innovation Science*, 1(1), 1–8.
- Zhang, J., Zhang, M., Yu, Y., & Yu, R. (2024). An innovative method integrating run theory and DBSCAN for complete three-dimensional drought structures. *Science of the Total Environment*, 926.