JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 13 No. 2, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i2.6231

PENERAPAN METODE K-MEANS CLUSTERING DALAM PEMETAAN KEMISKINAN KABUPATEN/KOTA DI INDONESIA UNTUK PERENCANAAN KEBIJAKAN YANG TEPAT

Mita Amelia[1*], Ahmad Faqih[2], Ade Rizki Rinaldi[3]

[1,2,3]STMIK IKMI Cirebon; Jl. Perjuangan No. 10B, Majasem, Cirebon, Jawa Barat 45135, Telp. (0231)490480

Received: 16 Februari 2025
Accepted: 19 Maret 2025
Published: 14 April 2025

Keywords:
Clustering, Kemiskinan, Kebijakan Sosial, *K-means*, *Knowledge Discovery in Database*

Corespondent Email:
miitaamelia888@gmail.com

Abstrak. Kemiskinan merupakan tantangan serius yang memerlukan pendekatan strategis berbasis data untuk mendukung kebijakan yang tepat sasaran. Penelitian ini bertujuan untuk mengelompokkan Kabupaten/Kota di Indonesia berdasarkan tingkat kemiskinan menggunakan algoritma K-Means dan mengevaluasi hasil pengelompokan untuk memberikan rekomendasi kebijakan. Metode analisis yang digunakan adalah Knowledge Discovery in Database (KDD) Process, yang melibatkan seleksi data, pra-pemrosesan, transformasi, penambangan data, dan evaluasi. Hasil analisis menunjukkan bahwa nilai k terbaik adalah 2 dengan nilai Davies-Bouldin Index sebesar 0,101. Klaster pertama (Cluster 0) mencakup wilayah dengan persentase penduduk miskin lebih rendah, rata-rata lama sekolah lebih tinggi, serta kondisi sosial ekonomi yang lebih baik dibandingkan klaster kedua. Sebaliknya, klaster kedua (Cluster 1) menunjukkan wilayah dengan tingkat kemiskinan signifikan, pendidikan rendah, dan minim infrastruktur dasar. Hasil penelitian ini menunjukkan bahwa pengelompokan menggunakan algoritma K-Means mampu mengidentifikasi wilayah prioritas untuk penanganan kemiskinan. Visualisasi klaster dan analisis karakteristik wilayah dapat mendukung perumusan kebijakan yang lebih efektif, terutama dalam peningkatan pendidikan, kesehatan, dan pengembangan infrastruktur.

Abstract. *Poverty is a significant challenge that requires a strategic, data-driven approach to support targeted policymaking. This study aims to cluster regencies/cities in Indonesia based on poverty levels using the K-Means algorithm and to evaluate the clustering results to provide policy recommendations. The analysis method applied is the Knowledge Discovery in Database (KDD) Process, which includes data selection, preprocessing, transformation, data mining, and evaluation. The analysis results indicate that the optimal k value is 2, with a Davies-Bouldin Index score of 0.101. The first cluster (Cluster 0) comprises regions with a lower percentage of impoverished populations, higher average years of schooling, and relatively better socioeconomic conditions compared to the second cluster. In contrast, the second cluster (Cluster 1) represents regions with significantly high poverty levels, lower education attainment, and inadequate basic infrastructure. This study demonstrates that the K-Means algorithm effectively identifies priority areas for poverty alleviation. Cluster visualization and regional characteristic analysis offer valuable insights to support the formulation of more effective policies, particularly in improving education, healthcare, and infrastructure development.*

1. PENDAHULUAN

Kemiskian merupakan masalah yang ada pada semua negara di dunia[1]. Kemiskinan menjadi salah satu permasalahan kompleks yang banyak ditemukan di negara-negara berkembang[2], Indonesia menjadi salah satu negara yang menghadapi tantangan besar terkait kemiskinan, masalah kemiskinan ini merupakan masalah kompleks yang memerlukan perhatian dan waktu lebih untuk mengatasinya.[3].Peningkatan populasi di Indonesia yang mencapai 278,8 juta jiwa pada tahun 2023 dipandang menguntungkan oleh sebagian orang, karena dapat mendukung pembangunan dan perekonomian, terutama jika tersedia banyak tenaga kerja. Namun, di sisi lain, pertumbuhan penduduk yang pesat juga dapat berkontribusi pada tingginya angka kemiskinan.[4].

Menurut UU No. 24/2004, kemiskinan merupakan kondisi sosio-ekonomi individu maupun kelompok yang kesulitan dalam mencukupi hak-hak fundamental mereka untuk menjaga dan meningkatkan kesejahteraan hidup yang optimal. Hak-hak fundamental tersebut meliputi akses pangan, kesehatan, pendidikan, pekerjaan, perumahan, air bersih, tanah, sumber daya alam dan lingkungan yang aman serta kesempatan terlibat dalam aktivitas sosial dan politik[5].

Isu sosial yang kian menjadi perhatian utama di Indonesia salah satunya merupakan kemiskinan[6]. Walaupun berbagai macam upaya telah dilakukan pemerintah untuk menekan angka kemiskinan, masalah kemiskinan di berbagai kabupaten dan kota masih tetap menjadi tantangan yang signifikan [7]. Berbagai faktor, seperti tingkat kemiskinan, pendidikan, pengeluaran per kapita, akses terhadap air bersih, dan perkembangan ekonomi yang diukur melalui PDRB, memiliki pengaruh yang besar terhadap kesejahteraan masyarakat di setiap kabupaten dan kota. Dengan menganalisis hubungan antara faktor-faktor ini, langkah-langkah yang lebih efektif dapat diambil untuk mengurangi kesenjangan antar wilayah.

Meskipun pemerintah telah menerapkan berbagai kebijakan untuk menurunkan tingkat kemiskinan, hasilnya belum merata di seluruh Indonesia. Beberapa kabupaten dan kota masih menghadapi tingkat kemiskinan yang tinggi, meskipun PDRB dan indikator sosial-ekonomi lainnya menunjukkan peningkatan [6]. Kondisi ini menimbulkan pertanyaan mengenai hubungan antara pertumbuhan ekonomi dan distribusi kesejahteraan sosial. Di sisi lain, kabupaten dan kota dengan PDRB serta indikator sosial-ekonomi yang rendah cenderung menghadapi tantangan besar dalam menurunkan angka kemiskinan. Selain itu, ketidakmerataan akses terhadap pendidikan, layanan kesehatan, dan infrastruktur juga berkontribusi pada masalah kemiskinan di berbagai daerah. Oleh karena itu, pengelompokan kabupaten dan kota dilakukan berdasarkan variabel sosial-ekonomi.

Beberapa penelitian sebelumnya telah mengkaji pengelompokan data untuk penyaluran bantuan sosial. Penelitian oleh Mayasari, dkk [8] membahas mengenai pengelompokan Kabupaten/Kota di Jawa Tengah berdasarkan data kemiskinan. Menggunakan K-means Cluster Analysis, data mencakup jumlah penduduk miskin, pertumbuhan penduduk, dan tingkat pengangguran pada tahun 2022. Hasilnya menunjukkan tiga kluster: 12 Kabupaten/Kota kategori kemiskinan rendah, 7 kategori tinggi, dan 16 kategori sedang.

Penelitian oleh Sukarno, dkk [9] berjudul membahas mengenai penentuan daerah prioritas untuk penanganan kemiskinan. Menggunakan data 2021-2023, penelitian ini menerapkan algoritma K-Means untuk mengelompokkan wilayah berdasarkan persentase penduduk miskin dan tingkat pengangguran. Proses clustering didukung oleh Silhouette Coefficient dan divisualisasikan melalui QGIS. Hasilnya menghasilkan dua kluster, rendah dan tinggi, dengan nilai Davies-Bouldin Index 0,600 dan skor Silhouette 0,550.

Pada penelitian lain oleh Sapitri, dkk [10] membahas mengenai pengelompokan Kabupaten/Kota di Jawa Barat berdasarkan tingkat kemiskinan menggunakan algoritma K-means, yang bertujuan mendukung pemerintah dalam perencanaan kebijakan dan penyaluran bantuan sosial. Dengan memanfaatkan pendekatan Knowledge Discovery in Databases (KDD) untuk menemukan pola-pola data yang valid, penelitian ini menggunakan perangkat lunak RapidMiner untuk mengolah data secara efektif.

Tujuan dari penelitian ini adalah untuk mengelompokkan kabupaten/kota di Indonesia

berdasarkan tingkat kemiskinan dengan menggunakan metode *K-means Clustering*. Metode ini diharapkan dapat memberikan pemahaman yang lebih baik mengenai pola kemiskinan dan kinerja ekonomi di berbagai daerah. Hasil dari penelitian ini diharapkan dapat memberikan informasi yang bermanfaat untuk mengoptimalkan kebijakan yang lebih efektif dalam menangani masalah kemiskinan. Selain itu, penelitian ini akan menggunakan variable sosial-ekonomi yang memengaruhi kemiskinan di Indonesia. Temuan ini dapat menjadi dasar bagi penelitian lebih lanjut mengenai faktor-faktor yang memengaruhi kemiskinan di Indonesia.

Penelitian ini menggunakan metode *K-means clustering* dengan pendekatan *KDD* (*Knowledge Discovery in Database*). Data yang digunakan berasal dari *Kaggle*, yang mencakup informasi tentang tingkat kemiskinan di kabupaten/kota di Indonesia. Metode *K-means* clustering diterapkan untuk mengelompokkan daerah berdasarkan karakteristik sosial-ekonomi yang serupa. Analisis dilakukan menggunakan perangkat lunak *RapidMiner* untuk memperoleh klaster yang optimal. Hasil klasterisasi akan digunakan untuk memahami pola kemiskinan dan ekonomi di Indonesia

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah teknik yang digunakan untuk menggali informasi berharga dari suatu basis data. Proses ini melibatkan penerapan berbagai metode canggih, seperti kecerdasan buatan, machine learning, dan teknik statistik, dengan tujuan untuk mengekstrak dan mengidentifikasi data yang relevan dari kumpulan data yang massif [11].

2.2. Algoritma K-Means

Algoritma K-Means adalah salah satu metode pengelompokan yang termasuk dalam kategori pembelajaran tanpa pengawasan (unsupervised learning), digunakan untuk mengelompokkan data ke dalam beberapa kelompok melalui sistem partisi [12]. Unsupervised learning adalah jenis algoritma machine learning yang digunakan untuk menarik kesimpulan dari dataset tanpa label respons. Metode yang paling umum dalam unsupervised learning adalah analisis kluster, yang berfungsi untuk mengidentifikasi pola tersembunyi atau pengelompokan dalam data[13].

Langkah-langkah dalam algoritma K-means dimulai dengan menentukan jumlah K yang diinginkan, yaitu jumlah kluster. Selanjutnya, titik pusat atau centroid untuk kluster tersebut ditentukan secara acak. Setelah itu, pada tahap iterasi, rumus berikut digunakan:

$$V_{ij} = \frac{1}{n_i}\sum_{k=0}^{n_i} X\,Kj \quad (1)$$

Keterangan:

V_{ij} = titik pusat rata-rata cluster ke i untuk varibel ke j

n_i =Jumlah anggota cluster ke i

i,k = indeks dari cluster

j = indeks dari variable

Xkj = nilai data ke k variabel j pada cluster tersebut

Sedangkan pada tahap hitung jarak antara tiap objek dengan centroid dengan rumus sebagai berikut:

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

Keterangan:

De = Jarak antara objek dengan centroid

I = banyaknya objek

(x,y) = koordinat objek

(s,t) = koordinat centroid

Setelah melakukan penghitungan jarak antara objek dengan centroid maka kelompokkan objek berdasarkan jarak centroid terdekat. Lakukan iterasi sehingga centroid bernilai optimal.

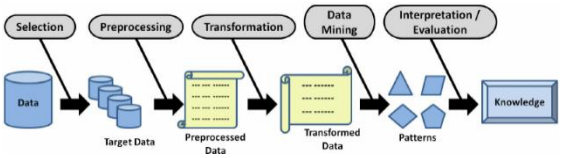
2.3. Clustering

Clustering adalah metode untuk mengelompokkan data berdasarkan kesamaan karakteristik yang dimilikinya[13]. Metode ini membagi data menjadi beberapa kelompok, di mana kesamaan antar data dalam satu kelompok lebih besar dibandingkan dengan kelompok lainnya. Konsep ini sederhana dan sejalan dengan cara berpikir manusia, terutama saat

menghadapi data besar yang perlu diringkas ke dalam kategori untuk memudahkan analisis. Selain itu, banyak data memiliki sifat alami yang cenderung membentuk kelompok tertentu [14].

3. METODE PENELITIAN

Metode penelitian ini menggunakan KDD (Knowledge Discovery in Databases) untuk memperoleh pengetahuan dari basis data. KDD akan diterapkan untuk menganalisis data kemiskinan dari dataset di Kaggle, guna mengidentifikasi pola dan hubungan antar variabel. Hasil analisis ini akan digunakan untuk mengelompokkan kabupaten/kota di Indonesia berdasarkan tingkat kemiskinan, dengan tahapan proses KDD dijelaskan pada **Gambar 1.**



Gambar 1. Tahapan KDD

3.1. Data Selection

Pada tahap data *selection*, dilakukan pengumpulan data operasional dari dataset yang tersedia di *Kaggle*. Pemilihan data dilakukan untuk memastikan bahwa hanya informasi yang relevan yang akan digunakan dalam analisis, dan hasil dari proses pemilihan ini akan disimpan terpisah dari basis data utama.

3.2. Preprocessing Data

Proses pre-processing bertujuan untuk menghilangkan baris yang mengandung nilai null (NaN) dan menjamin konsistensi data. Pada tahap ini, normalisasi variabel dilakukan serta penghapusan data yang tidak relevan untuk meningkatkan kualitas dataset. Langkah ini sangat krusial untuk memastikan bahwa analisis selanjutnya menggunakan data yang berkualitas dan sesuai dengan format yang diharapkan.

3.3. Transformation Data

Pada tahap transformasi, data diproses untuk memastikan hasil klasterisasi yang optimal. Semua variabel dinormalisasi agar berada pada skala yang konsisten. Normalisasi ini sangat penting agar algoritma K-means tidak terpengaruh oleh perbedaan skala antar variabel, sehingga hasil pengelompokan menjadi lebih akurat dan dapat dianalisis dengan efektif.

3.4. Data Mining

Pada tahap eksplorasi data, algoritma K-means Clustering akan digunakan untuk mengelompokkan kabupaten/kota berdasarkan tingkat kemiskinan dan faktor sosial-ekonomi lainnya. Proses ini akan menghasilkan kategori yang mencerminkan karakteristik masing-masing kelompok.

3.5. Evaluation

Langkah terakhir adalah merangkum dan menyimpulkan analisis terkait omset penjualan tahu dengan menerapkan metode regresi linear menggunakan RapidMiner, sekaligus memberikan gambaran mengenai prediksi omset penjualan di masa depan.

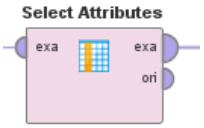
4. HASIL DAN PEMBAHASAN

Penelitian ini mengikuti tahapan yang sesuai dengan metodologi KDD (Knowledge Discovery in Databases).

4.1. Selection

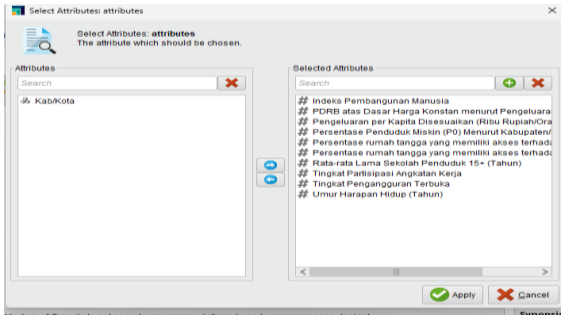
Dataset dalam penelitian ini diambil dari Kaggle dan terdiri dari 514 data tentang tingkat kemiskinan di berbagai kabupaten/kota di Indonesia. Atribut utama mencakup Persentase Penduduk Miskin (P0), Rata-rata Lama Sekolah Penduduk 15+, Pengeluaran per Kapita, Indeks Pembangunan Manusia, Umur Harapan Hidup, akses terhadap sanitasi dan air minum layak, Tingkat Pengangguran Terbuka, Tingkat Partisipasi Angkatan Kerja, dan PDRB berdasarkan pengeluaran. Dataset ini diolah menggunakan algoritma K-means Clustering untuk mengelompokkan wilayah berdasarkan tingkat kemiskinan dan karakteristik sosial-ekonomi lainnya.

Pada tahap Data Selection, digunakan operator Select Attributes untuk memilih atribut tertentu yang akan dianalisis. Operator ini memungkinkan penyaringan data, sehingga hanya variabel relevan yang disertakan dalam proses pengelompokan. **Gambar 2** menunjukkan tampilan operator Select Attributes yang digunakan dalam pemilihan atribut.



Gambar 2. Operator Select Atribut

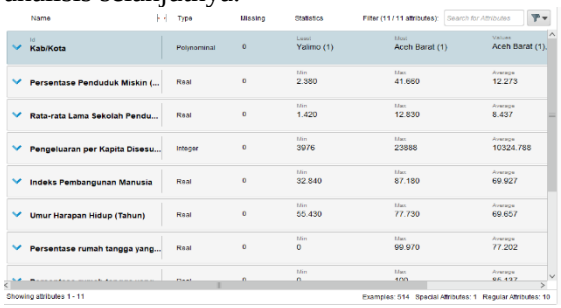
Hasil dari pemilihan atribut menggunakan operator *Select Attributes* ditampilkan pada **Gambar 3.** Gambar ini menunjukkan variabel-variabel yang telah dipilih dan akan digunakan dalam tahap analisis selanjutnya.



Gambar 3. Hasil Select Atribut

4.1.2. *Pre-processing*

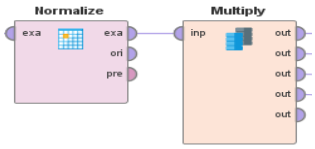
Pre-processing atau pembersihan data adalah tahap awal analisis, di mana data yang dipilih disiapkan dan dibersihkan untuk memastikan kualitas. Tahap ini penting untuk menghilangkan noise dan ketidakkonsistenan yang dapat memengaruhi hasil analisis. Dalam dataset ini, tidak ditemukan missing values, sehingga proses pembersihan tidak memerlukan operator tambahan. Hasil pemeriksaan pada **Gambar 4**. menunjukkan bahwa tidak ada data yang hilang, sehingga data siap untuk langkah analisis selanjutnya.



Gambar 4. Hasil Preprocessing

4.1.3. *Transformation*

Transformasi adalah tahap dalam analisis data yang mengubah data agar sesuai dengan format, struktur, dan skala yang diperlukan untuk analisis selanjutnya. Tahap ini penting untuk memastikan data dapat diolah secara optimal dalam algoritma yang dipilih. Dalam penelitian ini, penulis menggunakan operator Normalize untuk menstandarisasi nilai data dan operator Multiply untuk menyesuaikan skala variabel yang relevan serta mencari nilai K terbaik. Transformasi ini bertujuan meningkatkan akurasi dan konsistensi hasil analisis. **Gambar 5.** menunjukkan tampilan operator Normalize dan Multiply, sementara **Gambar 6.** memperlihatkan data setelah penerapan kedua operator tersebut.



Gambar 5. Operator Normalyze & Multiply

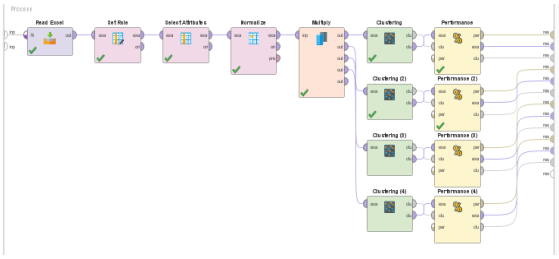
Transformasi data menggunakan operator Normalize menstandarkan nilai setiap atribut agar berada pada skala yang konsisten, sedangkan operator Multiply digunakan untuk menyesuaikan skala variabel tertentu sesuai tujuan analisis. Proses ini bertujuan mendukung pembentukan klaster yang lebih akurat pada tahap penggalian data berikutnya.

Gambar 6. Hasil Transformation

4.1.4. Data Mining

Berikut adalah desain tahapan pengelompokan data kemiskinan di Indonesia, yang menggambarkan seluruh proses mulai dari pengumpulan data, pembersihan, transformasi, hingga penerapan metode clustering menggunakan *K-means*. Desain ini bertujuan untuk memberikan gambaran yang jelas tentang langkah-langkah yang diambil dalam analisis data, memastikan setiap tahapan dilakukan secara sistematis dan efisien untuk menghasilkan pengelompokan yang akurat berdasarkan tingkat kemiskinan dan variabel terkait lainnya. **Gambar 7.** menunjukan desain tahapan pengelompokan data kemiskinan dari

K=2 sampai K5=5yang digunakan dalam penelitian ini.



Gambar 7. Proses K-Means Clustering

Operator utama dalam proses ini adalah K-means Clustering, yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiskinan dan variabel terkait. K-means membagi data menjadi beberapa klaster, di mana setiap data dalam satu klaster memiliki kesamaan yang lebih tinggi dibandingkan dengan klaster lainnya. Proses ini membantu mengidentifikasi pola atau karakteristik kemiskinan di berbagai daerah di Indonesia.

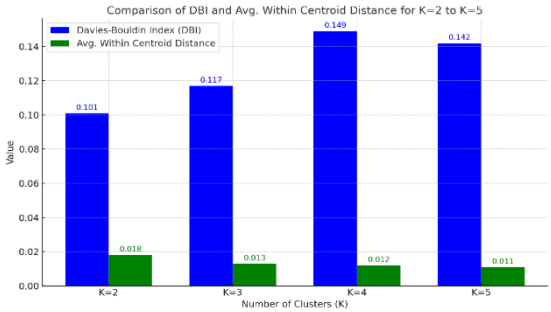
.

4.1.5. *Evaluation*

Hasil klasterisasi dalam penelitian ini dievaluasi menggunakan Davies-Bouldin Index (DBI), yang menilai kualitas klaster berdasarkan pemisahan antar klaster dan homogenitas dalam klaster. Semakin kecil nilai DBI, semakin baik kualitas klaster, menunjukkan pemisahan yang baik dan keseragaman tinggi. Hasilnya ditampilkan pada **Gambar 8.**

K	Cluster	DBI	Nilai Avg. Within Centroid Distance
K2	Cluster 0: 470 items	0.101	0.018
	Cluster 1: 44 items		
	Total number of items: 514		
K3	Cluster 0: 349 items	0.117	0.013
	Cluster 1: 131 items		
	Cluster 2: 34 items		
	Total number of items: 514		
K4	Cluster 0: 104 items	0.149	0.012
	Cluster 1: 243 items		
	Cluster 2: 22 items		
	Cluster 3: 145 items		
	Total number of items: 514		
K5	Cluster 0: 232 items	0.142	0.011
	Cluster 1: 120 items		
	Cluster 2: 22 items		
	Cluster 3: 13 items		
	Cluster 4: 127 items		
	Total number of items: 514		

Gambar 8. Hasil Pencarian K Optimal



Gambar 9. Hasil Diagram Batang Antar Cluster

Dari gambar diatas ditemukan bahwa K=2 merupakan nilai K yang optimal. Hal ini dikarenakan cluster dengan K=2 memiliki nilai DBI paling rendah mendekati nol, yang menunjukkan kualitas klaster yang baik. Hasil pengukuran ini dengan max runs sebanyak 10 dan Bregman Divergences menghasilkan nilai DBI sebesar 0.101.

Diagram pada **Gambar 8** membandingkan dua metrik evaluasi klasterisasi K-means: Davies-Bouldin Index (DBI) dan Average Within Centroid Distance untuk jumlah klaster bervariasi (K=2 hingga K=5). DBI menilai kualitas klaster berdasarkan pemisahan dan homogenitas, dengan nilai lebih kecil menunjukkan hasil lebih baik. Sementara itu, Average Within Centroid Distance mengukur jarak rata-rata titik data dalam klaster terhadap centroid-nya, dengan nilai lebih kecil menandakan keseragaman tinggi. Nilai DBI (batang biru) meningkat dengan bertambahnya jumlah klaster, menunjukkan kualitas yang kurang optimal. Sebaliknya, Average Within Centroid Distance (batang hijau) menurun, mendukung pemilihan K=2 sebagai jumlah klaster optimal dengan nilai DBI terendah (0.101).

Titik centroid dari setiap klaster dapat dilihat pada **Gambar 10.** yang menunjukkan posisi pusat setiap klaster dalam ruang data. Titik-titik centroid ini berfungsi sebagai titik representatif untuk masing-masing klaster, membantu dalam visualisasi dan interpretasi hasil pengelompokan.

K	Cluster	Rata-rata Lama sekolah (Tahun)	Pengeluaran per Kapita	IP	Umur Harapan Hidup (Tahun)	PRT dengan Sanitasi Layak	PRT dengan Air Minum Layak	TPT	Tingkat Partisipasi Angkatan	PD
2	Cluster 0	0.2128	0.6696	0.3376	0.7039	0.657	0.8072	0.8762	0.3992	0.0
	Cluster 1	0.636	0.39	0.1185	0.4534	0.4346	0.3989	0.5856	0.1565	0.0
3	Cluster 0	0.2512	0.1293	0.2844	0.6586	0.609	0.7696	0.8372	0.3153	0.0
	Cluster 1	0.5861	0.7616	0.4702	0.8154	0.7698	0.8955	0.9629	0.6078	0.1
	Cluster 2	0.7314	0.3459	0.0896	0.4156	0.4274	0.3245	0.5665	0.1431	0.0
4	Cluster 0	0.122	0.1862	0.4886	0.8268	0.7671	0.8945	0.9673	0.6594	0.1
	Cluster 1	0.7807	0.6041	0.3246	0.6942	0.6813	0.8313	0.8585	0.333	0.0
	Cluster 2	0.2659	0.5673	0.0764	0.3603	0.4104	0.2269	0.5322	0.1203	0.0
	Cluster 3	0.7858	0.374	0.2242	0.6083	0.5074	0.6684	0.8046	0.2921	0.0
5	Cluster 0	0.1875	0.1451	0.3112	0.6826	0.6712	0.8209	0.8426	0.3091	0.0
	Cluster 1	0.5925	0.7486	0.4412	0.7999	0.7478	0.8786	0.9594	0.6078	0.0
	Cluster 2	0.7858	0.0723	0.6691	0.8956	0.8352	0.8997	0.9857	0.745	0.5
	Cluster 3	0.2659	0.8372	0.2234	0.6053	0.4927	0.6644	0.8068	0.2955	0.0
	Cluster 4	0.7858	0.3963	0.2234	0.6053	0.4927	0.6644	0.8068	0.2955	0.0

Gambar 10 Titik Centroid

Selanjutnya, pada **Gambar 11.** menampilkan hasil pemodelan *K-means* yang dilakukan menggunakan perangkat lunak RapidMiner. Gambar tersebut menggambarkan distribusi data ke dalam klaster-klaster yang telah dibentuk berdasarkan algoritma *K-means*.

Gambar 11. Hasil Visualisasi K=2

Visualisasi ini menggambarkan pola pengelompokan data, termasuk posisi centroid untuk setiap klaster dalam 11 kelompok. **Cluster 0** menunjukkan persentase penduduk miskin yang lebih rendah, dengan rata-rata lama sekolah tinggi, pengeluaran per kapita rendah, dan Indeks Pembangunan Manusia (IPM) sedang. Umur harapan hidupnya tinggi, meskipun PDRB rendah, menandakan ekonomi yang lemah. Akses air minum layak tinggi, tetapi sanitasi masih rendah. Sebaliknya, **Cluster 1** memiliki persentase penduduk miskin yang sangat tinggi, mencerminkan tingkat kemiskinan signifikan. Rata-rata lama sekolah dan pengeluaran per kapita rendah, menunjukkan akses pendidikan dan daya beli yang terbatas, serta IPM dan umur harapan hidup yang rendah. PDRB lebih rendah dibandingkan cluster 0, dengan akses air minum layak dan sanitasi yang sangat terbatas.

5. KESIMPULAN

Penelitian ini berhasil mengelompokkan kabupaten/kota di Indonesia menggunakan metode K-means Clustering, yang membagi wilayah menjadi dua klaster: Cluster_0, yang mencakup daerah dengan tingkat kemiskinan rendah, dan Cluster_1, yang mencakup daerah dengan tingkat kemiskinan lebih tinggi, sehingga menunjukkan perbedaan signifikan antar daerah. Evaluasi menggunakan Davies-Bouldin Index (DBI) menghasilkan nilai 0.101 untuk K=2, yang menunjukkan kualitas klaster yang baik dan menjadikan jumlah klaster ini optimal untuk mendukung kebijakan yang lebih efektif. Oleh karena itu, rekomendasi untuk optimasi kebijakan meliputi fokus pada pengembangan sektor ekonomi dan infrastruktur di Cluster_0, serta peningkatan pendidikan, pelatihan keterampilan, dan pemberdayaan ekonomi di Cluster_1 untuk secara efektif mengurangi kemiskinan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

[1] L. G. Rady Putra and A. Anggrawan, “Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 205–214, 2021, doi: 10.30812/matrik.v21i1.1554.

[2] E. Mansi, E. Hysa, M. Panait, and M. C. Voica, “Poverty-A challenge for economic development? Evidences from Western Balkan countries and the European union,” *Sustain.*, vol. 12, no. 18, 2020, doi: 10.3390/SU12187754.

[3] R. Nabila, I. P. Syahna, K. I. Daulay, and S. Wulandari, “Kemiskinan dan Ketimpangan Melalui Prevelensi Ketidakcukupan Konsumsi Masyarakat di Indonesia,” *El-Mujtama J. Pengabdi. Masy.*, vol. 3, no. 3, pp. 760–769, 2023, doi: 10.47467/elmujtama.v3i3.2934.

[4] P. Sari, E. Efan, and R. Syahri, “Analisis Clustering Data Penduduk Miskin Menggunakan Algoritma K-Means,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 2194–2199, 2024, doi: 10.36040/jati.v8i2.9433.

[5] F. Febriansyah and S. Muntari, “Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam,”

JISKA (Jurnal Inform. Sunan Kalijaga), vol. 8, no. 1, pp. 66–77, 2023, doi: 10.14421/jiska.2023.8.1.66-77.

[6] T. A. Munandar, "Penerapan Algoritma Clustering Untuk Pengelompokan Tingkat Kemiskinan Provinsi Banten," *JSiI (Jurnal Sist. Informasi)*, vol. 9, no. 2, pp. 109–114, 2022, doi: 10.30656/jsii.v9i2.5099.

[7] B. Baskoro, A. Gunaryati, and A. Rubhasy, "Klasifikasi Penduduk Kurang Mampu Dengan Metode K-Means untuk Optimalisasi Program Bantuan Sosial," *Infomatek*, vol. 25, no. 1, pp. 41–48, 2023, doi: 10.23969/infomatek.v25i1.7271.

[8] S. N. Mayasari and J. Nugraha, "Implementasi K-Means Cluster Analysis untuk Mengelompokkan Kabupaten / Kota Berdasarkan Data Kemiskinan di Provinsi Jawa Tengah Tahun 2022," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 317–329, 2023.

[9] N. Sukarno Wijaya, M. Jajuli, and B. Arif Dermawan, "Penerapan Algoritma K-Means Clustering Dalam Menentukan Daerah Prioritas Penanganan Kemiskinan Di Wilayah Jawa Timur," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp. 7579–7584, 2024, doi: 10.36040/jati.v8i4.10248.

[10] S. Sapitri, R. Astuti, and F. M Basysyar, "Pengelompokkan Jumlah Penduduk Miskin Di Jawa Barat Menggunakan Algoritma K-Means," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 2895–2900, 2024, doi: 10.36040/jati.v8i3.9600.

[11] R. Nugraha, N. Suarna, I. Ali, and D. Rohman, "OPTIMASI PENGELOLAAN SAMPAH MELALUI MODEL PENGELOMPOKAN DENGAN ALGORITMA K-MEANS," vol. 13, no. 1, pp. 646–652, 2025.

[12] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/JTSI

[13] N. Buslim and R. P. Iswara, "Pengembangan Algoritma Unsupervised Learning Technique Pada Big Data Analysis di Media Sosial sebagai media promosi Online Bagi Masyarakat," *J. Tek. Inform.*, vol. 12, no. 1, pp. 79–96, 2019, doi: 10.15408/jti.v12i1.11342.