



KLASTERISASI TINGKAT KESEJAHTERAAN MASYARAKAT PADA PROVINSI DI INDONESIA MENGGUNAKAN ALGORITMA K-MEANS

Marlisa Koliham¹, Anita Diana², Rizky Tahara Shita^{3*}

^{1,2,3} Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta

^{1,2,3} Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, Jakarta 12260

e-mail : marlisakoliham@gmail.com¹, anita.diana@budiluhur.ac.id², rizky.taharashita@budiluhur.ac.id^{3*}

ABSTRAK

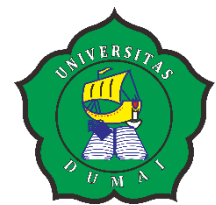
Pembangunan nasional di Indonesia bertujuan untuk meningkatkan kesejahteraan masyarakat secara merata. Namun, ketimpangan kesejahteraan antarprovinsi masih menjadi tantangan utama. Penelitian ini bertujuan untuk mengelompokkan 38 provinsi di Indonesia berdasarkan tingkat kesejahteraan menggunakan empat indikator utama: Upah Minimum Provinsi (UMP), Indeks Pembangunan Manusia (IPM), persentase penduduk miskin, serta tingkat Pengangguran Terbuka (TPT). Pendekatan yang digunakan adalah metode data mining dengan algoritma K-Means Clustering yang dianalisis melalui tahapan CRISP-DM. Data sekunder tahun 2021–2024 diperoleh dari Badan Pusat Statistik dan Kementerian Ketenagakerjaan. Penentuan jumlah grup yang paling sesuai dilakukan melalui Metode Elbow, dan pemeriksaan keabsahan dari hasil pengelompokan menggunakan Davies Bouldin Index (DBI). Hasil penelitian menunjukkan lima klaster kesejahteraan, yaitu: sangat tinggi (3 provinsi), menengah tinggi (8 provinsi), menengah (13 provinsi), rendah (11 provinsi), dan sangat rendah (5 provinsi), dengan nilai DBI sebesar 0,845 yang mengindikasikan kualitas klaster yang baik. Temuan ini diharapkan dapat mendukung pengambilan kebijakan yang lebih terarah dalam upaya pemerataan kesejahteraan di Indonesia.

Kata Kunci : *Kesejahteraan Provinsi, K-Means, Data Mining, CRISP-DM, Elbow Method*

ABSTRACT

National development in Indonesia aims to improve the overall and equitable welfare of society. However, disparities in welfare levels among provinces remain a significant challenge. This study seeks to cluster 38 provinces in Indonesia based on welfare levels using four key indicators: Provincial Minimum Wage (UMP), Human Development Index (HDI), poverty rate, and open unemployment rate. A data mining approach was employed using the K-Means clustering algorithm, following the CRISP-DM methodology. Secondary data from 2021 to 2024 were obtained from the Central Bureau of Statistics (BPS) and the Ministry of Manpower. The optimal number of clusters was determined using the Elbow Method, while the clustering results were validated using the Davies Bouldin Index (DBI). The findings reveal five distinct welfare clusters: very high (3 provinces), upper-middle (8 provinces), middle (13 provinces), low (11 provinces), and very low (5 provinces), with a DBI value of 0.845 indicating good clustering quality. These results are expected to support more data-driven and targeted policymaking efforts by both central and regional governments to promote equitable development

Key Word : *Provincial Welfare, K-Means, Data Mining, CRISP-DM, Elbow Method*



1. PENDAHULUAN

Pemerataan kesejahteraan menjadi salah satu tujuan utama dalam pelaksanaan pembangunan nasional di Indonesia. Namun, kenyataannya menunjukkan bahwa ketidakmerataan antarprovinsi masih merupakan masalah serius. Mengacu pada data Badan Pusat Statistik (BPS) tahun 2023, terdapat perbedaan signifikan dalam nilai Indeks Pembangunan Manusia (IPM) serta tingkat kemiskinan antara provinsi seperti DKI Jakarta dengan Papua, yang mencerminkan adanya kesenjangan pembangunan antarwilayah (BPS, 2024). Kesenjangan ini juga terlihat dari variasi Upah Minimum Provinsi (UMP) serta tingkat pengangguran terbuka yang tidak merata. Keempat indikator UMP, IPM, kemiskinan, dan pengangguran terbuka memiliki hubungan yang kuat dalam menggambarkan kompleksitas kondisi kesejahteraan masyarakat.

Walaupun data dari lembaga resmi seperti BPS dan Kementerian Ketenagakerjaan sudah tersedia secara menyeluruh, penggunaannya dalam analisis kesejahteraan masyarakat masih kurang maksimal. Sedikit penelitian telah mencoba mengintegrasikan keempat indikator ini dalam satu model analisis yang terpadu. Akibatnya, potensi data yang ada belum dimanfaatkan dengan baik untuk mendukung pengambilan keputusan berbasis data dalam perencanaan pembangunan daerah. Oleh sebab itu, perlu pendekatan analisis yang mampu mengelompokkan provinsi-provinsi di Indonesia berdasarkan persamaan karakteristik kesejahteraan mereka, sehingga kebijakan yang lebih tepat dan efektif dapat disusun.

Salah satu pendekatan yang dapat diterapkan adalah teknik klusterisasi dalam *data mining*, utamanya melalui algoritma *K-Means*. Algoritma ini memiliki kapasitas untuk mengelompokkan data ke dalam beberapa kluster berdasarkan kesamaan atribut, sehingga dapat memetakan provinsi-provinsi dengan kondisi kesejahteraan serupa secara objektif dan efisien. Beberapa penelitian sebelumnya telah memperlihatkan keefektifan *K-Means* dalam pengelompokan wilayah berdasarkan indikator sosial ekonomi (Sihombing & Sihombing, 2022). Namun, penerapan metode ini untuk analisis tingkat kesejahteraan masyarakat secara menyeluruh di Indonesia masih jarang dilakukan.

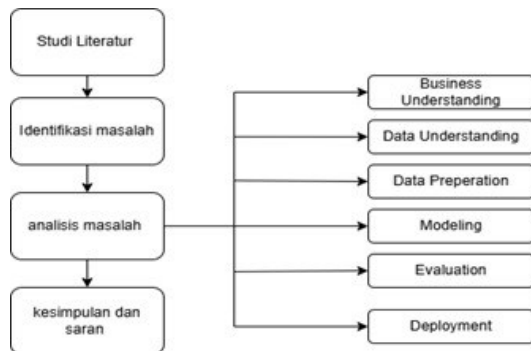
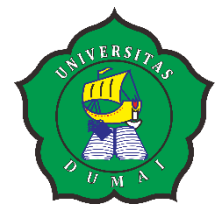
Penelitian sebelumnya mayoritas hanya menggunakan satu atau dua indikator saja. Misalnya, penelitian oleh Nurfathullah dan Purnamasari (2024) yang mengelompokkan provinsi berdasarkan tingkat pengangguran terbuka dalam periode 2018–2023 dengan *K-Means* dan *Elbow Method*. Sedangkan penelitian oleh Khalif et al. (n.d.) mengklasifikasikan provinsi berdasarkan tingkat kemiskinan dan pengeluaran per kapita. Padahal, indikator seperti

UMP, IPM, tingkat kemiskinan, dan tingkat pengangguran terbuka memberikan gambaran yang lebih komprehensif mengenai kondisi kesejahteraan. UMP mencerminkan kemampuan ekonomi minimum para pekerja, IPM menggambarkan kualitas hidup melalui pendidikan, kesehatan, dan daya beli (Rukhan Saputra & Zaida Muflih, 2025), sedangkan tingkat kemiskinan dan pengangguran menunjukkan tekanan ekonomi serta tantangan di bidang ketenagakerjaan suatu wilayah (Murharni & Andriyanto, 2022).

Berbeda dengan penelitian yang telah dilakukan sebelumnya, kajian ini menggabungkan keempat indikator utama tersebut untuk mengelompokkan 38 provinsi di Indonesia dengan menggunakan algoritma *K-Means*. Pendekatan ini diharapkan mampu memberikan pandangan yang lebih holistik terhadap kondisi kesejahteraan di berbagai wilayah. Selain itu, model divalidasi menggunakan *Davies Bouldin Index (DBI)* untuk menjamin kualitas hasil klusterisasi. Tujuan dari studi ini adalah untuk mengkategorikan provinsi-provinsi yang ada di Indonesia. Berdasarkan indikator kesejahteraan menggunakan algoritma *K-Means* serta menetapkan jumlah kluster optimal menggunakan *Elbow Method*. Hasil penelitian diharapkan dapat menjadi pertimbangan bagi pemerintah pusat dan daerah dalam merumuskan kebijakan yang berfokus pada pemerataan kesejahteraan masyarakat.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *data mining* dengan metode klusterisasi menggunakan algoritma *K-Means*. Proses penelitian dilakukan dengan mengacu pada kerangka kerja CRISP-DM (Cross-Industry Standard Process for Data Mining) yang terdiri dari enam tahapan, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. CRISP-DM merupakan standar pemrosesan data mining yang dirancang agar setiap tahap dalam proses tersebut dapat berjalan secara terorganisir dan efisien (Dzulhijjah et al., n.d.).



Gambar 1. Tahapan Penelitian

Pada gambar 1. dapat dilihat tahapan yang digunakan dalam penelitian dengan menggunakan CRISP-DM. Berikut penjelasannya.

a) Studi Literatur

Tahap ini bertujuan untuk mengumpulkan data dan sumber yang terkait dengan isu penelitian melalui berbagai media, baik dalam bentuk fisik maupun digital. Penelitian pustaka dilakukan untuk memperkuat landasan teori serta mendukung kerangka pemikiran dalam studi ini.

b) Identifikasi Masalah

Langkah ini dilakukan untuk mengidentifikasi permasalahan yang relevan dengan topik penelitian, berdasarkan hasil kajian literatur, observasi langsung, serta sumber data yang mendukung pelaksanaan penelitian.

c) Analisa Masalah

Analisis dilakukan untuk memperoleh gambaran menyeluruh tentang permasalahan yang diteliti. Pada bagian ini, digunakan pendekatan metode CRISP-DM yang terdiri dari enam tahapan utama, yaitu:

1. *Business Understanding*

Pada tahap ini dilakukan identifikasi masalah, tujuan penelitian, dan manfaat dari proses klusterisasi. Fokus utama dari Penelitian ini bertujuan untuk melakukan klasifikasi terhadap provinsi-provinsi di Indonesia.

2. *Data Understanding*

Data yang digunakan merupakan data sekunder dari tahun 2021 hingga 2024 yang diperoleh dari Badan Pusat Statistik (BPS) dan Kementerian Ketenagakerjaan. Indikator yang digunakan terdiri dari empat variabel utama:

- Upah Minimum Provinsi (UMP)
- Indeks Pembangunan Manusia (IPM)
- Tingkat Kemiskinan
- tingkat Pengangguran Terbuka (TPT)

3. *Data Preperation*

Tahapan ini mencakup proses *data cleaning*, transformasi data, dan normalisasi menggunakan metode Min-Max. Tujuannya adalah untuk menyamakan skala data agar proses klusterisasi menjadi lebih akurat. Normalisasi dilakukan dengan rumus:

$$v' = \frac{v - \min A}{\max A - \min A} (\text{new}_{\max A} - \text{new}_{\min A}) + \text{new}_{\min A}$$

Keterangan:

v = nilai hasil normalisasi

v = nilai asli

min A = nilai minimum dari atribut A

max A = nilai maksimum dari atribut A

new_{maxA} = rentang jarak terbesar

new_{MinA} = rentang jarak terkecil.

4. *Evaluation*

Pada tahap ini penelitian mengevaluasi hasil model dengan menggunakan metode Elbow. Evaluasi dilakukan dengan mengamati hasil kluster dan menilai logika pembagian kelompok. Validasi dilakukan dengan melihat pemisahan antarprovinsi berdasarkan karakteristik kesejahteraan dan melihat apakah proses klusterisasi sudah berjalan sesuai tujuan penelitian.

5. *Deployment*

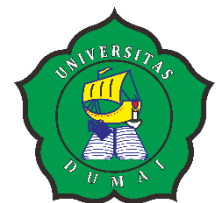
Tahapan terakhir di mana hasil penelitian disajikan dalam bentuk laporan, tabel, dan visualisasi agar mudah dipahami oleh pihak yang membutuhkan. Penyajian hasil ini diharapkan dapat memberikan informasi yang bermanfaat bagi pengambilan Keputusan terkait upah minimum di tiap provinsi.

d) Kesimpulan dan Saran

Terdapat Hasil analisis masalah dengan metode CRISP-DM menghasilkan kesimpulan dan rekomendasi untuk proses secara keseluruhan.

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) dan Kementerian Ketenagakerjaan Republik Indonesia (satudata.kemnaker.go.id). Data terdiri dari 38 provinsi dan mencakup empat indikator utama kesejahteraan, yaitu **Upah Minimum Provinsi (UMP), Indeks Pembangunan Manusia (IPM), Tingkat kemiskinan, serta tingkat Pengangguran Terbuka (TPT)**. dengan cakupan tahun 2021 hingga 2024.



2.2 Data Pra-processing

Sebelum dilakukan proses klusterisasi, data melalui tahapan *data cleaning* dan *data transformation*. Data yang kosong, duplikat, atau tidak relevan dihapus. Selain itu, dilakukan juga penyatuan format data dari beberapa sumber agar seragam. Proses ini memastikan data bersih dan siap digunakan dalam proses analisis.

2.3 Pemilihan Dataset

Indikator yang dipilih merupakan representasi dari dimensi ekonomi, sosial, dan ketenagakerjaan dalam mengukur kesejahteraan. Pemilihan variabel didasarkan pada relevansi data dan ketersediaan yang konsisten antarprovinsi dalam periode waktu yang sama.

2.4 Normalisasi Data

Setelah proses pembersihan data selesai, semua variabel dinyatakan dalam bentuk yang distandarisasi dengan menggunakan metode Normalisasi Min-Max untuk membuatnya berada dalam kisaran 0 sampai 1. Proses normalisasi ini sangat krusial karena algoritma K-Means mengandalkan jarak Euclidean yang memiliki kepekaan tinggi terhadap skala data.

2.5 Algoritma K-Means

Merupakan metode klusterisasi non-hierarki yang populer dalam data mining. Teknik ini bekerja dengan membagi data ke dalam sejumlah kluster berdasarkan kemiripan, menggunakan prinsip *partitioning* yang dipengaruhi oleh posisi pusat kluster (centroid). Berbeda dengan klusterisasi hierarki yang menggabungkan item secara bertahap, K-Means melakukan iterasi untuk meminimalkan variasi dalam kluster dan memaksimalkan perbedaan antar kluster (Rizki Nugroho & Edo Hendrawan, 2022). Data akan dikelompokkan ke dalam beberapa kluster, di mana anggota dalam satu kluster memiliki karakteristik yang mirip, namun berbeda dengan kluster lainnya (Dhewayani et al., n.d.).

2.6 Euclidean Distance

Euclidean Distance merupakan metode yang digunakan untuk mengukur jarak antara dua titik dalam ruang Euclidean, dengan dasar perhitungan yang mengacu pada hubungan geometris antara koordinat kedua titik tersebut. Secara matematis, perhitungan ini mirip dengan rumus *Pythagoras* yang digunakan untuk mengukur jarak antara dua titik dalam satu atau lebih dimensi.

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

Keterangan:

De = Jarak Euclidean antara objek ke centroid

i = Indeks objek ke- i

$x_i y_i$ = Koordinat objek ke- i pada dimensi x dan y

$s_i t_i$ = Koordinat centroid pada dimensi x dan y

2.7 Metode Elbow

Metode Elbow merupakan teknik yang digunakan untuk menentukan jumlah kluster optimal dalam algoritma K-Means. Metode ini bekerja dengan menghitung total jarak antara data dan pusat kluster (centroid) untuk berbagai nilai K , lalu memplot hasilnya dalam grafik. Titik “tekukan” atau *elbow* pada grafik menunjukkan jumlah kluster terbaik karena menggambarkan saat penurunan jarak mulai melambat secara signifikan (Muningsih & Kiswati, 2018). Menurut (Mahendrasyah, 2023), metode ini memberikan gambaran tentang perubahan persentase hasil klusterisasi seiring penambahan jumlah kluster. Nilai-nilai tersebut divisualisasikan dalam grafik yang menjadi dasar pengambilan keputusan jumlah kluster yang paling representatif.

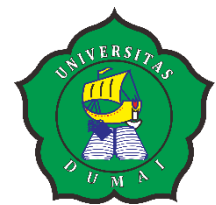
3. HASIL DAN PEMBAHASAN

Bagian ini membahas hasil dari proses klusterisasi provinsi di Indonesia berdasarkan indikator kesejahteraan, serta evaluasi dan interpretasi hasil tersebut. Proses klusterisasi menggunakan algoritma K-Means dengan pendekatan metode Elbow untuk menentukan jumlah kluster optimal dan validasi model dengan Davies Bouldin Index (DBI).

3.1 Penentuan Jumlah Kluster Optimal

1. Pemodelan dengan Algoritma K-Means

Langkah pertama dalam proses pembuatan model dilakukan dengan menggunakan algoritma *K-Means*. Proses pemodelan ini memanfaatkan perangkat lunak *data mining* RapidMiner untuk memperoleh nilai *Davies-Bouldin Index (DBI)* pada berbagai variasi jumlah kluster. Tujuan dari pemodelan awal ini adalah untuk menemukan jumlah kluster yang paling ideal berdasarkan nilai DBI terendah.



Tabel 1. Pemodelan dengan Algoritma K-Means

Penentuan Jumlah Kluster Terbaik	Hasil Davies Bouldin Index
2 Kluster	1,007
3 Kluster	1,099
4 Kluster	0,905
5 Kluster	0,845

Berdasarkan hasil yang ditampilkan pada tabel di atas, nilai DBI terendah diperoleh pada jumlah kluster sebanyak lima, yaitu sebesar 0,845. Hal ini menunjukkan bahwa semakin rendah nilai DBI, semakin baik kualitas kluster yang dihasilkan. Oleh karena itu, ditetapkan bahwa jumlah kluster terbaik untuk pemodelan adalah 5 kluster.

2. Pemodelan dengan Algoritma K-Medoids

Selain menggunakan algoritma K-means, dilakukan juga pemodelan menggunakan algoritma klusterisasi lainnya yaitu K-Medoids sebagai pembanding. Pemodelan dilakukan dengan jumlah kluster bervariasi dari 2 hingga 5.

Gambar 2. Nilai Davies Bouldin Index K-Medoids

Penentuan Jumlah Kluster Terbaik	Hasil Davies Bouldin Index
2 Kluster	1,007
3 Kluster	1,099
4 Kluster	0,905
5 Kluster	0,845

Berdasarkan informasi yang terdapat dalam tabel di atas, nilai terkecil untuk *Davies-Bouldin Index (DBI)* diperoleh ketika terdapat empat kluster, yaitu sebesar 1,574. Angka ini mengindikasikan bahwa kualitas kluster yang dihasilkan oleh algoritma *K-Medoids* tidak lebih unggul dibandingkan dengan *K-Means*, meskipun masih menunjukkan performa yang cukup baik dalam proses pengelompokan data.

Gambar di bawah ini menampilkan hasil terbaik yang diperoleh berdasarkan nilai *Davies-Bouldin Index*, yaitu menggunakan algoritma *K-Means* dengan nilai DBI sebesar 0,845. Nilai tersebut dianggap baik karena merupakan nilai terendah yang dicapai. Semakin kecil nilai DBI atau semakin mendekati nol, maka semakin tinggi tingkat optimalitas kluster yang dihasilkan.

Tabel 3. hasil terbaik berdasarkan nilai Davies Bouldin Index,

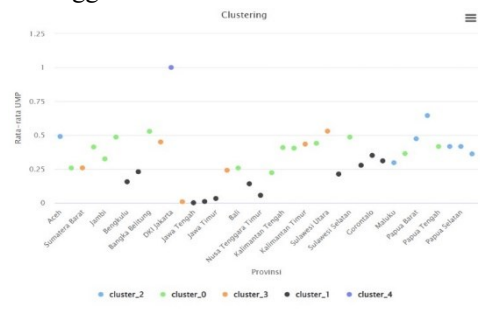
Algoritma	Hasil Davies Bouldin Index
K-Means	0,845
K-Medoids	1,574

3.2 Hasil Klusterisasi dan Visualisasi

Visualisasi data menggunakan scatter plot untuk setiap indikator yang digunakan dalam penelitian, yaitu Upah Minimum Provinsi (UMP), Indeks Pembangunan Manusia (IPM), Tingkat Kemiskinan, dan Tingkat Pengangguran Terbuka (TPT). Visualisasi scatter plot ini bertujuan untuk melihat distribusi nilai dari masing-masing indikator terhadap seluruh provinsi di Indonesia.

1. Data UMP

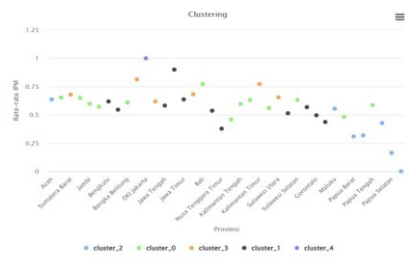
Kluster 0 mencakup provinsi dengan UMP menengah seperti Sumatera Utara, Kalimantan Selatan, dan Sulawesi Selatan. Kluster 1 didominasi oleh provinsi dengan UMP rendah, seperti Jawa Tengah, NTB, dan Sulawesi Barat. Kluster 2 berisi wilayah timur seperti Papua dan Maluku dengan UMP yang cenderung tinggi, meskipun persebarannya tidak merata. Kluster 3 terdiri dari provinsi dengan UMP tinggi seperti Jawa Barat, Kalimantan Timur, dan Sulawesi Utara. Sementara itu, Kluster 4 hanya ditempati oleh DKI Jakarta sebagai provinsi dengan UMP tertinggi secara nasional.



Gambar 2. Data UMP

2. Data IPM

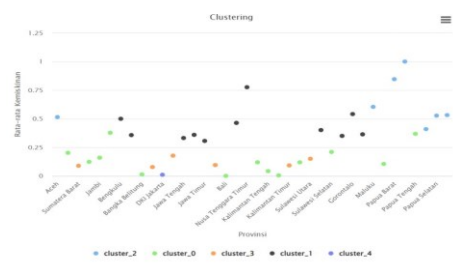
Kluster 0 terdiri dari provinsi dengan IPM cukup tinggi seperti Jawa Barat, Banten, dan Kalimantan Timur. Kluster 1 mencakup wilayah dengan IPM menengah seperti Jawa Tengah, NTB, dan Sulawesi Tenggara. Kluster 2 didominasi oleh provinsi dibagian wilayah timur Indonesia seperti Maluku, Papua dan Aceh dengan IPM rendah. Kluster 3 berisi provinsi dengan IPM tinggi yang juga muncul di Kluster 0, seperti Sulawesi Utara dan Kepulauan Riau. Sementara itu, Kluster 4 hanya ditempati oleh DKI Jakarta sebagai provinsi dengan IPM tertinggi secara nasional.



Gambar 3. Data IPM

3. Data Tingkat Kemiskinan

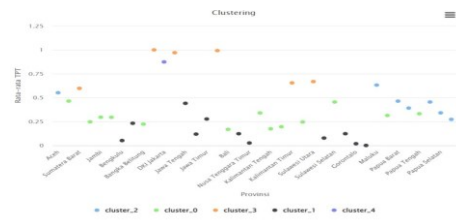
Klaster 0 mencakup provinsi dengan tingkat kemiskinan menengah seperti Sumatera Utara, Kalimantan Selatan, dan Bali. Klaster 1 berisi wilayah dengan kemiskinan cukup tinggi, antara lain Jawa Tengah, Nusa Tenggara Barat, dan Sulawesi Tenggara. Klaster 2 didominasi oleh provinsi di kawasan timur seperti Papua, Maluku, dan Aceh, yang memiliki tingkat kemiskinan tertinggi. Klaster 3 mencakup provinsi dengan tingkat kemiskinan rendah seperti Jawa Barat, Kalimantan Timur, dan Kepulauan Riau. Sedangkan Klaster 4 hanya diisi oleh DKI Jakarta, sebagai provinsi dengan tingkat kemiskinan terendah secara nasional.



Gambar 4. Data Tingkat Kemiskinan

4. Data Tingkat Pengangguran Terbuka

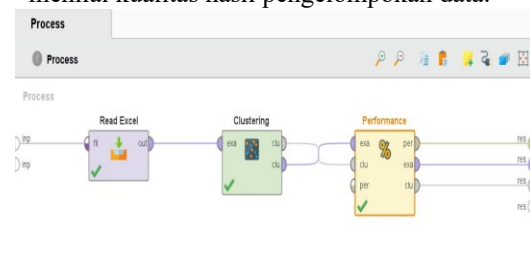
Klaster 0 terdiri dari provinsi dengan tingkat kemiskinan menengah seperti Sumatera Utara, Kalimantan Selatan, dan Bali. Klaster 1 mencakup wilayah dengan tingkat kemiskinan cukup tinggi, seperti Jawa Tengah, NTT, dan Sulawesi Tengah. Klaster 2 didominasi oleh provinsi di kawasan timur seperti Papua, Papua Barat, dan Aceh yang memiliki tingkat kemiskinan tertinggi. Klaster 3 berisi provinsi dengan kemiskinan rendah seperti Jawa Barat, Kalimantan Timur, dan Kepulauan Riau. Sementara itu, Klaster 4 hanya ditempati oleh DKI Jakarta sebagai provinsi dengan tingkat kemiskinan terendah secara nasional.



Gambar 5. Data TPT

3.3 Penyajian Model Terbaik

Tahap ini menyajikan output akhir dari pemodelan terbaik yang telah dijalankan menggunakan algoritma K-Means dengan jumlah klaster sebanyak 5. Pemrosesan dilakukan melalui aplikasi RapidMiner dengan memanfaatkan sejumlah operator utama, seperti K-Means dan Performance (Clustering), untuk menilai kualitas hasil pengelompokan data.



Gambar 6. Penyajian Model Terbaik

Gambar di atas merupakan alur kerja proses yang digunakan. Operator K-Means digunakan untuk membentuk klaster berdasarkan kemiripan data, dan operator Performance digunakan untuk mengukur seberapa baik hasil klaster terbentuk berdasarkan *average within centroid distance* dan *Davies Bouldin Index*.

Cluster Model

```
Cluster 0: 13 items
Cluster 1: 11 items
Cluster 2: 7 items
Cluster 3: 6 items
Cluster 4: 1 items
Total number of items: 38
```

Gambar 7. Cluster Model

Dari hasil proses klasterisasi, didapatkan pembagian sebanyak 5 klaster dari total 38 provinsi. Rinciannya adalah, klaster_0 terdiri dari 13 provinsi klaster_1 terdiri dari 11 provinsi klaster_2 terdiri dari 7 provinsi klaster_3 terdiri dari 6 provinsi dan klaster_4 terdiri dari 1 provinsi.



3.4 Perhitungan dengan Algoritma Terpilih

Pada tahap ini, dilakukan proses perhitungan manual menggunakan algoritma K-Means yang telah dipilih. Pengolahan data dilakukan setelah melalui tahap transformasi, sehingga data siap untuk dianalisis dengan metode klusterisasi K-Means. Perhitungan menggunakan data sampel sebanyak 7 entri.

jumlah kluster (k) dan nilai *average within centroid distance*. Grafik tersebut memperlihatkan

Taembl 4. Perhitungan dengan Algoritma Terpilih

Provinsi	UMP	IPM	KEMISKINAN	TPT
Aceh	0,490	0,636	0,514	0,551
Sumatera Utara	0,257	0,653	0,202	0,464
Sumatera Barat	0,258	0,679	0,089	0,597
Riau	0,412	0,650	0,123	0,247
Jambi	0,324	0,598	0,160	0,294
Sumatera Selatan	0,485	0,572	0,375	0,294
Bengkulu	0,155	0,620	0,500	0,051

Menentukan centroid awal, tahap ini akan dipilih titik centroid secara acak. Pusat kluster awal dapat dilihat pada tabel di bawah ini:

Tabel 5. Centroid awal

Centroid	UMP	IPM	KEMISKINAN	TPT
C0	0,440	0,559	0,119	0,245
C1	0,416	0,427	0,410	0,453
C2	0,000	0,581	0,332	0,441
C3	1,000	1,000	0,010	0,874
C4	0,310	0,437	0,364	0,000

Jarak data 1 dengan centroid 1

$$= \sqrt{(0,490 - 0,440)^2 + (0,636 - 0,559)^2 + (0,514 - 0,199)^2 + (0,551 - 0,245)^2}$$

$$= 0,509$$

Jarak data 1 dengan centroid 2

$$= \sqrt{(0,490 - 0,416)^2 + (0,636 - 0,427)^2 + (0,514 - 0,410)^2 + (0,551 - 0,453)^2}$$

$$= 0,264$$

Jarak data 1 dengan centroid 3

$$= \sqrt{(0,490 - 0,000)^2 + (0,636 - 0,581)^2 + (0,514 - 0,332)^2 + (0,551 - 0,441)^2}$$

$$= 0,538$$

Jarak data 1 dengan centroid 4

$$= \sqrt{(0,490 - 1,000)^2 + (0,636 - 1,000)^2 + (0,514 - 0,010)^2 + (0,551 - 0,874)^2}$$

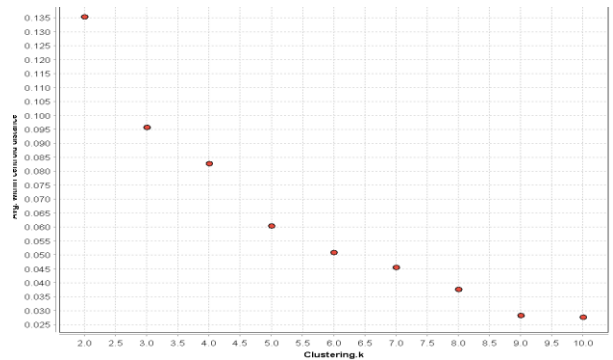
$$= 0,866$$

Jarak data 1 dengan centroid 5

$$= \sqrt{(0,490 - 0,310)^2 + (0,636 - 0,437)^2 + (0,514 - 0,364)^2 + (0,551 - 0,000)^2}$$

$$= 0,632$$

Gambar 8. Jarak data 1 dengan Centroid 1



Gambar 9. Grafik Scatter

Jarak data 1 dengan centroid 5

$$= \sqrt{(0,490 - 0,310)^2 + (0,636 - 0,437)^2 + (0,514 - 0,364)^2 + (0,551 - 0,000)^2}$$

$$= 0,632$$

Selanjutnya menghitung jarak data 7 dengan centroid 1

$$= \sqrt{(0,155 - 0,440)^2 + (0,620 - 0,559)^2 + (0,500 - 0,119)^2 + (0,051 - 0,245)^2}$$

$$= 0,517$$

Jarak data 7 dengan centroid 2

$$= \sqrt{(0,155 - 0,416)^2 + (0,620 - 0,427)^2 + (0,500 - 0,410)^2 + (0,051 - 0,453)^2}$$

$$= 0,524$$

Jarak data 7 dengan centroid 3

$$= \sqrt{(0,155 - 0,000)^2 + (0,620 - 0,581)^2 + (0,500 - 0,332)^2 + (0,051 - 0,441)^2}$$

$$= 0,453$$

Jarak data 7 dengan centroid 4

$$= \sqrt{(0,155 - 1,000)^2 + (0,620 - 0,437)^2 + (0,500 - 0,010)^2 + (0,051 - 0,874)^2}$$

$$= 1,332$$

Jarak data 7 dengan centroid 5

$$= \sqrt{(0,155 - 0,310)^2 + (0,620 - 0,559)^2 + (0,500 - 0,364)^2 + (0,051 - 0,000)^2}$$

$$= 0,280$$

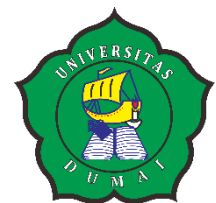
Gambar 10. Jarak data 1 dengan Centroid 1

Hasil perhitungan data nilai centroid adalah anggota dari setiap kluster yang jaraknya terpendek dari pusat kluster ditunjukkan pada Tabel

Tabel 6. Hasil Perhitungan

Provinsi	C0	C1	C2	C3	C4	Jarak Terdekat
Aceh	0,509	0,264	0,538	0,866	0,632	0,264
Sumatera Utara	0,312	0,346	0,298	0,936	0,540	0,298
Sumatera Barat	0,415	0,461	0,399	0,858	0,702	0,399
Riau	0,095	0,417	0,505	0,935	0,418	0,095
Jambi	0,138	0,354	0,395	0,988	0,393	0,138
Sumatera Selatan	0,266	0,228	0,509	0,958	0,369	0,228
Bengkulu	0,517	0,524	0,453	1,332	0,280	0,280

Dari hasil analisis terhadap setiap data dalam kluster, dapat diketahui kluster mana yang menjadi tempat yang sesuai bagi masing-masing data. Proses pengelompokan dilakukan dengan mempertimbangkan nilai jarak terendah dari setiap data terhadap pusat kluster. Oleh karena itu, data berhasil terintegrasi ke dalam kluster yang sesuai berdasarkan hasil perhitungan jarak tersebut.

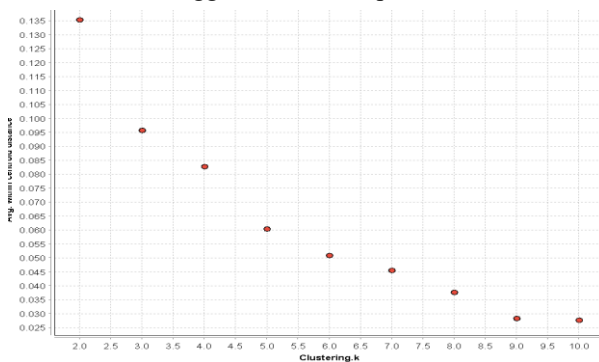


Gambar 15. Hasil Klaster

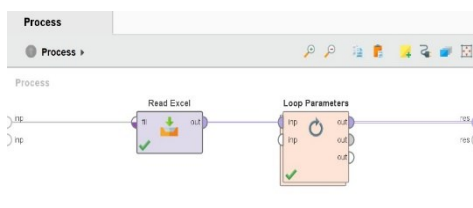
Provinsi	UMP	IPM	KEMISKINAN	TPT	KLASTER
Aceh	0,490	0,636	0,514	0,551	2
Sumatera Utara	0,257	0,653	0,202	0,464	3
Sumatera Barat	0,258	0,679	0,089	0,597	3
Riau	0,412	0,650	0,123	0,247	1
Jambi	0,324	0,598	0,160	0,294	1
Sumatera Selatan	0,485	0,572	0,375	0,294	2
Bengkulu	0,155	0,620	0,500	0,051	5

3.5 Pengujian

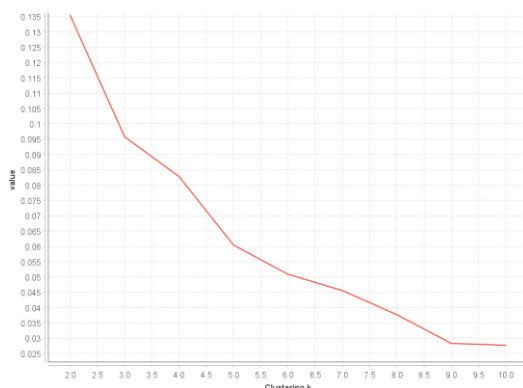
Pada bagian ini dijelaskan cara mengukur performa model yang terbaik. Metode pengujian yang digunakan dalam penelitian ini adalah Elbow method. Berikut adalah proses pengujian yang dilakukan menggunakan alat Rapidminer:



Gambar 11. Grafik Scatter



Gambar 12. Pengujian



Gambar 13. Hasil penerapan Metode Elbow

Visualisasi pada gambar di atas menampilkan hasil penerapan metode Elbow untuk menentukan jumlah klaster yang paling tepat dalam proses pemodelan. Grafik menunjukkan hubungan antara jumlah klaster (klastering.k) dan nilai rata-rata jarak terhadap pusat klaster (*Average Within*

Centroid Distance). Terlihat adanya penurunan signifikan pada nilai tersebut mulai dari K = 3 ke K = 4, dan berlanjut hingga K = 5. Pada titik K = 5, grafik membentuk sudut menyerupai “siku”, yang mengindikasikan titik optimal sebelum kurva mulai datar. Hal ini menunjukkan bahwa K = 5 merupakan pilihan jumlah klaster terbaik karena pada titik tersebut terjadi penurunan performa yang paling mencolok. Dengan demikian, jumlah klaster optimal dapat diidentifikasi pada K = 5.

Grafik Scatter ini ditambahkan untuk memperjelas visualisasi hasil klasterisasi, sehingga hubungan antara jumlah klaster (*Clustering.k*) dan nilai rata-rata jarak terhadap pusat klaster (*Average Within Centroid Distance*) dapat diinterpretasikan dengan lebih baik. Dari grafik terlihat bahwa terjadi penurunan signifikan dalam nilai rata-rata jarak mulai dari K = 2 hingga K = 5. Titik K = 5 ini membentuk pola “siku” (elbow) yang menjadi indikator bahwa lima klaster merupakan jumlah yang paling optimal.

4. KESIMPULAN

Penelitian ini berhasil mengelompokkan 38 provinsi di Indonesia ke dalam klaster-klaster menggunakan algoritma K-Means berdasarkan 4 indikator utama, yaitu Upah Minimum Provinsi (UMP), Indeks Pembangunan Manusia (IPM), tingkat kemiskinan, serta tingkat Pengangguran Terbuka (TPT) berdasarkan data periode 2021 hingga 2024.

Jumlah klaster optimal yang didapat dari hasil proses klasterisasi adalah 5 klaster. Hal ini ditentukan melalui metode Elbow, yang menunjukkan titik siku (elbow point) pada K = 5, serta diperkuat dengan interpretasi hasil klasterisasi yang membagi provinsi-provinsi di Indonesia ke dalam 5 kelompok kesejahteraan yang berbeda.

Klaster 0 Kesejahteraan Rendah hingga Menengah (13 provinsi), Klaster 1 Kesejahteraan Menengah (11 provinsi), Klaster 2 Kesejahteraan Sangat Rendah (7 provinsi), Klaster 3 Kesejahteraan Menengah Tinggi (6 provinsi), Klaster 4 Kesejahteraan Tinggi (1 provinsi).

5. REFERENSI

Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., Jajuli, M., HSRonggo Waluyo, J., Telukjambe Timur, K., Karawang, K., & Barat, J. (n.d.). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*. <https://doi.org/10.34010/jati.v12i1>



Dzulhijjah, D. A., Bagus Herlambang, M., & Haifan, M. (n.d.). *IMPLEMENTASI FRAMEWORK CRISP-DM UNTUK PROSES DATA MINING APLIKASI CREDIT SCORING PT. XYZ.*

<https://doi.org/10.24198/jmi.v17.n2.35025.12>
7-135

Khalif, A., Nur Hasanah, A., Hafizh Ridwan, M., Nurina Sari, B., Kunci -Kemiskinan, K., Crispdm, M., & Berkelanjutan, P. (n.d.). Klasterisasi Tingkat Kemiskinan di Indonesia Menggunakan Algoritma K-Means. In *Generation Journal* (Vol. 8, Issue 1).

Mahendrasyah, I. (2023). *KLASTERISASI INDEKS STANDAR PENCEMARAN UDARA MENGGUNAKAN ALGORITMA K-MEANS.*

Muningsih, E., & Kiswati, S. (2018). SISTEM APLIKASI BERBASIS OPTIMASI METODE ELBOW UNTUK PENENTUAN CLUSTERING PELANGGAN. In *JOUTICA* (Vol. 3, Issue 1).

Murharni, S., & Andriyanto, S. (2022). 977d41f963a350636d1c4d2449f8acc8d18e. *Jurnal Informatika*, 22(TINGKAT PENGANGGURAN TERBUKA TAHUN 2016-2018 DAN 2019-2021).

Nurfathullah, M., & Purnamasari, I. (2024). IMPLEMENTASI K-MEANS UNTUK MENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN INDEKS JUMLAH PENGANGGURAN TERBUKA. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 2).

Rizki Nugroho, M., & Edo Hendrawan, I. (2022). *Penerapan Algoritma K-Means Untuk Klasterisasi Data Obat Pada Rumah Sakit ASRI.* 16(1).
<https://journal.uniku.ac.id/index.php/ilkom>

Rukhan Saputra, N., & Zaida Muflih, G. (2025). Pengelompokan Wilayah Indonesia Berdasarkan Komponen Indeks Pembangunan Manusia dengan Pendekatan Algoritma K-Means Clustering. *SKANIKA: Sistem Komputer Dan Teknik Informatika*, 8(1), 156–167.

Sihombing, S. C., & Sihombing, D. A. (2022). Pengelompokan Tingkat Kesejahteraan Masyarakat di Sumatera Utara dengan Metode K-Means Clustering. *Jurnal Matematika Integratif*, 17(2), 127.