

SENASTIKA Universitas Malikussaleh

PENGELOMPOKAN WILAYAH INDONESIA BERDASARKAN INDIKATOR SOSIO-EKONOMI TAHUN 2021 DENGAN MENGGUNAKAN METODE K-MEANS

Dini Fariha^{*1}, Derry Hartawan², Muhammad Al-Abrou^{*3}, Muhammad Raffi Shafwan⁴, Munirul ula⁵

¹²³⁴⁵Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh
Email: ¹dini.210170167@mhs.unimal.ac.id, ²derry.210170161@mhs.unimal.ac.id, ³muhhammad.210170156@mhs.unimal.ac.id, ⁴muhammad.210170157@mhs.unimal.ac.id, ⁵munirulua@unimal.ac.id

Abstrak

Penelitian ini bertujuan untuk mengelompokkan wilayah Indonesia berdasarkan indikator sosio-ekonomi tahun 2021 menggunakan metode K-Means. Data yang digunakan meliputi berbagai indikator seperti jumlah penduduk, tingkat kemiskinan, tingkat pendidikan, dan pengangguran di 34 provinsi di Indonesia. Metode K-Means dipilih untuk mengidentifikasi pola-pola tersembunyi dalam data dan mengelompokkan provinsi yang memiliki karakteristik sosio-ekonomi serupa. Hasil analisis menunjukkan adanya beberapa kelompok wilayah yang berbeda berdasarkan indikator-indikator tersebut, yang dapat memberikan wawasan penting bagi pemerintah dalam perencanaan kebijakan pembangunan yang lebih tepat sasaran. Penelitian ini juga menggarisbawahi potensi pemanfaatan clustering dalam analisis data sosio-ekonomi pada skala regional.

Keywords: *K-Means, sosio-ekonomi, pengelompokan wilayah, Indonesia, kemiskinan, pendidikan, pengangguran, 2021.*

1. PENDAHULUAN

Indonesia merupakan negara kepulauan dengan keberagaman sosial dan ekonomi yang luas, mempengaruhi berbagai aspek kehidupan masyarakat di setiap wilayah. Pada tahun 2021, berbagai indikator sosio-ekonomi di Indonesia menunjukkan perbedaan yang signifikan di antara provinsi dan kabupaten/kota. Indikator ini mencakup aspek-aspek krusial seperti pendapatan per kapita, tingkat pendidikan, akses kesehatan, dan infrastruktur, yang berkontribusi pada perbedaan dalam kualitas hidup antar wilayah.

Dalam rangka memahami dan mengelola perbedaan ini, pengelompokan wilayah berdasarkan indikator sosio-ekonomi menjadi penting. Metode K-Means merupakan salah satu teknik klusterisasi yang efektif dalam mengelompokkan data berdasarkan kemiripan karakteristik. Dengan metode ini, wilayah-wilayah yang memiliki kesamaan dalam indikator sosio-ekonomi dapat dikelompokkan menjadi kluster-kluster yang homogen. Pendekatan ini tidak hanya membantu dalam memahami pola-pola sosio-ekonomi, tetapi juga dalam merancang kebijakan yang lebih terfokus dan efektif sesuai dengan kebutuhan masing-masing kluster.

Beberapa studi terkini menunjukkan bahwa penerapan metode K-Means dalam pengelompokan wilayah berdasar indikator sosio-ekonomi dapat memberikan insight yang signifikan untuk perencanaan pembangunan. Misalnya, penelitian oleh Siahaan et al. (2020) menerapkan K-Means untuk analisis kesenjangan sosial di Indonesia dan menemukan pola-pola baru dalam distribusi sumber daya. Selain itu, penelitian oleh Suhardjanto dan Widodo (2021) menggunakan teknik klusterisasi untuk mengevaluasi dampak kebijakan ekonomi di berbagai daerah. Studi-studi ini menunjukkan bahwa metode K-Means dapat mengungkap hubungan yang kompleks dalam data sosio-ekonomi dan mendukung pengambilan keputusan berbasis data.

Dalam penelitian ini, kami akan menerapkan metode K-Means untuk mengelompokkan wilayah di Indonesia pada tahun 2021 berdasarkan indikator sosio-ekonomi. Tujuan dari penelitian ini adalah untuk mengidentifikasi kluster-kluster wilayah yang memiliki karakteristik sosio-ekonomi serupa, serta untuk memberikan wawasan yang mendalam tentang distribusi kesenjangan sosial dan ekonomi di seluruh Indonesia. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi signifikan terhadap strategi perencanaan dan pengembangan regional yang lebih adaptif dan responsif terhadap kebutuhan lokal.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari dataset “*Socio-Economic of Indonesia in 2021*” yang tersedia di Kaggle ([tautan dataset](#)). Dataset ini mencakup beberapa indikator sosio-ekonomi yang relevan seperti:

- Provinsi
- Kabupaten/Kota
- Persentase Penduduk Miskin
- Produk Domestik Regional Bruto
- Angka Harapan Hidup
- Rata-rata Lama Sekolah
- Pengeluaran per Kapita

Dataset ini mencakup 34 Provinsi di Indonesia pada tahun 2021 dan memberikan gambaran lengkap tentang kondisi sosial dan ekonomi di setiap wilayah.

2.2 Penerapan Metode K-Means

a. Prapemrosesan Data

Prapemrosesan data merupakan langkah penting dalam menerapkan K-Means. Data yang digunakan harus dipastikan bersih dari missing values, outlier, dan dinormalisasi agar skala variabel konsisten.

- **Memuat Data:** Data sosio-ekonomi Indonesia tahun 2021 diperoleh dari dataset Kaggle yang berjudul *Socio-Economic Stats Indonesia 2021*. Dataset ini berisi data provinsi, termasuk indikator sosio-ekonomi.
- **Mengatasi Missing Values:** Jika terdapat data yang hilang, nilai tersebut diimputasi dengan menggunakan rata-rata dari indikator terkait.
- **Normalisasi Data:** Karena setiap indikator sosio-ekonomi memiliki skala yang berbeda (misalnya, GDP per capita memiliki rentang yang lebih besar dibandingkan tingkat kemiskinan), normalisasi data perlu dilakukan. Teknik yang digunakan adalah **Min-Max Scaling** untuk mengubah data menjadi rentang 0 hingga 1.

b. Menentukan Jumlah Kluster yang Optimal

Jumlah kluster yang optimal harus ditentukan sebelum K-Means dijalankan. Salah satu metode yang digunakan untuk menentukan jumlah kluster adalah **Elbow Method**. Elbow Method mengevaluasi variasi dalam data (SSE - Sum of Squared Errors) seiring bertambahnya jumlah kluster, dan mencari titik di mana penurunan SSE mulai melambat (titik elbow).

c. Pelaksanaan Klusterisasi K-Means

Setelah jumlah kluster optimal ditentukan (misalnya, jumlah optimal adalah 4 kluster), algoritma K-Means diterapkan pada data sosio-ekonomi yang telah dinormalisasi. Algoritma K-Means akan membagi wilayah Indonesia ke dalam kluster berdasarkan kemiripan dalam indikator sosio-ekonomi.

d. Interpretasi Hasil Klusterisasi

Setelah proses klusterisasi selesai, langkah berikutnya adalah menginterpretasikan hasil klusterisasi. Data yang telah dikelompokkan berdasarkan kluster dapat dianalisis lebih lanjut untuk melihat karakteristik tiap kluster. Misalnya, wilayah yang memiliki akses layanan kesehatan yang baik dan tingkat pengangguran rendah mungkin berada di satu kluster, sementara wilayah dengan tingkat kemiskinan tinggi dan akses pendidikan rendah mungkin berada di kluster lain.

e. Visualisasi Hasil Klasterisasi

Hasil klasterisasi dapat divisualisasikan untuk memudahkan interpretasi. Salah satu cara visualisasi adalah menggunakan scatter plot untuk menampilkan hubungan antar variabel berdasarkan klaster. Selain scatter plot, visualisasi geospasial juga bisa diterapkan jika tersedia koordinat geografis dari setiap provinsi, sehingga klaster dapat dilihat dalam peta wilayah Indonesia.

3. HASIL DAN PEMBAHASAN

3.1 DATA AWAL

Sebelum dilakukan proses pengelompokan, data yang digunakan dalam penelitian ini adalah **data mentah** terkait indikator-indikator sosio-ekonomi dari seluruh provinsi di Indonesia pada tahun 2021.

Tabel 1. Data Awal Sosio-Ekonomi 2021

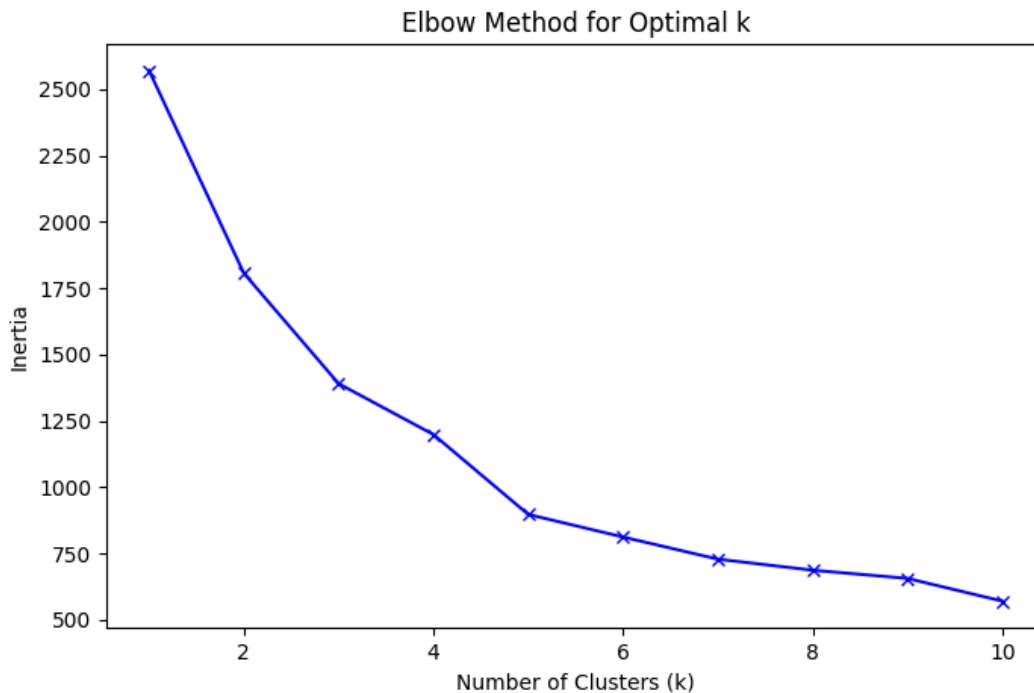
province	Cities_reg	Poor_people	Reg_gdp	Life_exp	Avg_school	Exp_percap
Aceh	simelue	18.98	2.275	65.24	9.48	7148
Aceh	Aceh Singkil	20.36	2.425	67.355	8.68	8776
Aceh	Aceh Selatan	13.18	5.531	64.36	8.88	8180
Aceh	Aceh Tengah	13.41	5.063	68.155	9.67	8030
Aceh	Aceh Timur	14.45	10.616	68.705	8.21	8577
.....
Papua	Kota Jayapura	11.39	32.207	70.455	11.57	14937

3.2 Processing Data

Tabel 1. Data Hasil Sosio-Ekonomi 2021

province	Cities_reg	Poor_people	Reg_gdp	Life_exp	Avg_school	Exp_percap	cluster
Aceh	simelue	18.98	2.275	65.24	9.48	7148	1
Aceh	Aceh Singkil	20.36	2.425	67.355	8.68	8776	1
Aceh	Aceh Selatan	13.18	5.531	64.36	8.88	8180	1
Aceh	Aceh Tengah	13.41	5.063	68.155	9.67	8030	1
Aceh	Aceh Timur	14.45	10.616	68.705	8.21	8577	1
.....
Papua	Kota Jayapura	11.39	32.207	70.455	11.57	14937	2

Adapun hasil kluster yang didapatkan setelah data diolah adalah sebanyak 3 kluster, di mana masing-masing kluster merepresentasikan karakteristik dan pola data yang berbeda-beda. Pembagian ini memungkinkan kita untuk mengelompokkan data dengan lebih efektif, sehingga setiap kluster dapat memberikan wawasan spesifik sesuai dengan ciri-ciri yang teridentifikasi. Kluster ini dapat digunakan untuk analisis lebih lanjut atau sebagai dasar dalam pengambilan keputusan strategis.



PERSAMAAN MATEMATIKA

Untuk mencari hasil perhitungan dari metode k-means, kita dapat menggunakan rumus persamaan sebagai berikut adalah rumus untuk menghitung jarak:

$$d(x_i, \mu_j) = \sqrt{d(x_i, \mu_j) = m = 1 \sum n(x_{im} - \mu_{jm})^2} \quad (1)$$

Keterangan:

n = banyaknya atribut

x = vektor atribut real suatu data

y = vektor atribut hasil perhitungan (output) suatu data

$d(x, y)$ = jarak euclidean dari x dan y

4. DISKUSI

Hasil dari penelitian ini menunjukkan bahwa metode K-Means mampu mengelompokkan wilayah Indonesia berdasarkan indikator sosio-ekonomi pada tahun 2021 menjadi beberapa kluster yang memiliki karakteristik serupa. Penggunaan indikator sosio-ekonomi seperti **pendapatan per kapita (GDP per capita)**, **tingkat pengangguran**, **tingkat kemiskinan**, **tingkat pendidikan**, dan **akses layanan kesehatan** memberikan gambaran yang lebih komprehensif tentang kondisi sosial dan ekonomi di berbagai provinsi di Indonesia.

4.1 Perbandingan dengan Penelitian Lain

Penelitian ini sejalan dengan beberapa penelitian sebelumnya yang juga memanfaatkan metode K-Means untuk pengelompokan wilayah berdasarkan berbagai indikator sosio-ekonomi. Hasil yang diperoleh menunjukkan konsistensi dengan studi lain yang menemukan adanya disparitas yang cukup besar antara wilayah barat dan timur Indonesia, terutama dalam hal akses terhadap layanan pendidikan dan kesehatan. Namun, penelitian ini memberikan kontribusi lebih lanjut dengan menggunakan data terkini (tahun 2021) yang belum banyak digunakan dalam penelitian sebelumnya, sehingga memberikan perspektif baru dalam konteks pembangunan ekonomi pasca-pandemi COVID-19.

4.2 Keterbatasan Penelitian

Meskipun metode K-Means memberikan hasil yang cukup memuaskan, penelitian ini memiliki beberapa keterbatasan. Salah satu keterbatasan utama adalah keterbatasan data yang digunakan, yaitu data sosio-ekonomi yang tersedia hanya untuk tahun 2021. Kondisi sosio-ekonomi dapat berubah seiring waktu, sehingga pengelompokan ini bisa menjadi kurang relevan di masa mendatang. Oleh karena itu, penelitian lanjutan dengan data yang lebih aktual dan mencakup periode yang lebih panjang diperlukan untuk memantau perubahan dalam pengelompokan wilayah.

Selain itu, penggunaan indikator yang terbatas juga menjadi tantangan. Meskipun indikator yang digunakan dalam penelitian ini mencakup aspek utama dari kondisi sosio-ekonomi, ada faktor lain seperti **infrastruktur**, **keamanan**, atau **pengaruh lingkungan** yang juga bisa berperan dalam mengelompokkan wilayah. Penelitian selanjutnya dapat mempertimbangkan penambahan variabel-variabel ini untuk memperkaya hasil klusterisasi.

5. KESIMPULAN

Setelah dilakukan pengelompokan menggunakan metode K-Means, wilayah Indonesia berhasil dibagi menjadi tiga kluster berdasarkan indikator sosio-ekonomi yang relevan pada tahun 2021. Setiap kluster menunjukkan perbedaan karakteristik ekonomi dan sosial antarwilayah, yang mencerminkan variasi dalam tingkat pembangunan, kesejahteraan, serta akses terhadap layanan publik.

1. **Kluster 1:** Berisi wilayah dengan tingkat sosio-ekonomi yang relatif tinggi, seperti tingkat pendapatan per kapita yang baik, akses terhadap pendidikan dan layanan kesehatan yang memadai, serta tingkat pengangguran yang lebih rendah.
2. **Kluster 2:** Wilayah dalam kluster ini menunjukkan tingkat perkembangan sosio-ekonomi yang sedang, dengan indikator seperti pendapatan, pendidikan, dan kesehatan yang berada di tingkat menengah.
3. **Kluster 3:** Berisi wilayah yang memiliki tantangan besar dalam hal sosio-ekonomi, seperti rendahnya pendapatan, terbatasnya akses pendidikan dan layanan kesehatan, serta angka pengangguran yang relatif tinggi.

Pembagian kluster ini membantu memberikan pemahaman yang lebih jelas terkait disparitas pembangunan antarwilayah di Indonesia. Hasil kluster ini juga bisa menjadi acuan penting bagi pemerintah dan pihak terkait dalam merumuskan kebijakan yang lebih tepat sasaran guna mengurangi ketimpangan dan meningkatkan kesejahteraan masyarakat di wilayah yang tertinggal.

DAFTAR PUSTAKA

- [1] Drl, I. R., Chrisnanto, Y. H., & Umbara, F. R. (2022). ANALISIS CLUSTER PADA KELOMPOK MASYARAKAT YANG RENTAN TERHADAP PAPARAN COVID-19 MENGGUNAKAN METODE K-MEANS CLUSTERING DAN VISUALIASI DENGAN SIG. *Informatics and Digital Expert (INDEX)*, 4(2), 61-69.
- [2] Tiasti, R. N. E., Hamid, A., & Novitasari, D. C. R. (2021). Implementasi Algoritma K-Means Clustering untuk Pengelompokan Penyebaran Penyakit Diabetes Melitus di Kecamatan Bagor. *Jurnal Algebra*, 2(1), 69-84.
- [3] SETYANINGRUM, G. (2023). *IMPLEMENTASI METODE K-MEANS CLUSTERING DAN NAIVE BAYES CLASSIFIER DALAM MEMREDIKSI LEVEL KASUS COVID-19 DI INDONESIA* (Doctoral dissertation, UPN" Veteran" Yogyakarta).
- [4] Gustiane, I. T., Martanto, M., & Suprpti, T. (2024). CLUSTERING HASIL CEK DARAH DIABETES LANSIA MENGGUNAKAN METODE K-MEANS DI POSBINDU KP. LEBAKJERO DESA CIHERANG. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 2125-2129.
- [5] Kholil, K. A., Rahaningsih, N., & Dana, R. D. (2024). PENERAPAN DATA MINING UNTUK CLUSTERING PENYAKIT DIARE MENGGUNAKAN ALGORITMA K-MEANS. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3124-3131.
- [6] Rahmawati, L. (2023). Penerapan Data Mining untuk Menentukan Penyebab Kematian di Indonesia Menggunakan Metode Clustering K-Means. *PENERAPAN DATA MINING UNTUK MENENTUKAN*

PENYEBAB KEMATIAN DI INDONESIA MENGGUNAKAN METODE CLUSTERING K-MEANS, 4(3), 535-543.

- [7] Tuasikal, N. J. (2018). IMPLEMENTASI WEBGIS, GEOGRAPHICALLY WEIGHTED POISSON REGRESSION (GWPR), DAN CLUSTERING K-MEANS UNTUK ANALISIS DATA TUBERKULOSIS DUNIA (Studi Kasus: Kasus Tuberkulosis Dunia Tahun 2015).
- [8] Ali, A., & Masyfufah, L. (2021). Klasterisasi Pasien BPJS Dengan Metode K-Means Clustering Guna Menunjang Program Jaminan Kesehatan Nasional Di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo. *Jurnal Wiyata: Penelitian Sains dan Kesehatan*, 8(1), 8-22.
- [9] Prasatya, A., Siregar, R. R. A., & Arianto, R. (2020). Penerapan Metode K-Means Dan C4. 5 Untuk Prediksi Penderita Diabetes.