
Aplikasi Segmentasi Negara Berdasarkan Indikator Ekonomi dan Kesehatan Berbasis Algoritma K-Means dan Streamlit

Srifunky Ledy Purba¹⁾, Jhon Travolta Tangdirerung²⁾, Ester Enjelina³⁾,
Muhammad Ifan Rifani Ihsan⁴⁾

¹⁻³ Informatika, Fakultas Teknik & Informatika, Universitas Bina Sarana Informatika

Email: srifunkypurba5@gmail.com, jhontravolta76@gmail.com, esterenjelina2004@gmail.com,
ifan.mii@bsi.ac.id

Abstrak

Penelitian ini memiliki tujuan untuk mengelompokkan negara-negara di dunia berdasarkan indikator ekonomi dan kesehatan menggunakan algoritma K-Means, serta menampilkannya melalui aplikasi interaktif Streamlit. Data berasal dari World Development Indicators dan mencakup variabel seperti PDB, harapan hidup, angka kelahiran, angka kematian, akses listrik, dan pengeluaran pemerintah untuk kesehatan. Data dianalisis melalui langkah-langkah pembersihan, standarisasi, dan penentuan jumlah kluster optimal menggunakan metode Elbow, yang menghasilkan tiga kluster utama: negara berpenghasilan tinggi, menengah, dan rendah. Visualisasi menggunakan Principal Component Analysis (PCA) menunjukkan pemisahan kluster yang jelas dan mencerminkan perbedaan tingkat kesejahteraan antar negara. Aplikasi berbasis Streamlit memudahkan pengguna untuk memahami pola pembangunan global secara interaktif. Hasil penelitian ini diharapkan dapat menjadi dasar bagi pengembangan kebijakan pembangunan yang efektif dan adil.

Kata Kunci: Kata Kunci: K-Means, Streamlit, PDB, Kesehatan, Analisis Kluster

Abstract

This study aims to classify countries around the world based on economic and health indicators using the K-Means algorithm, and to present them through an interactive Streamlit application. The data comes from the World Development Indicators and includes variables such as GDP, life expectancy, birth rate, death rate, access to electricity, and government health expenditure. The data is analyzed through steps of cleaning, standardization, and determining the optimal number of clusters using the Elbow method, resulting in three main clusters: high-income, middle-income, and low-income countries. Visualization using Principal Component Analysis (PCA) shows clear cluster separation and reflects differences in welfare levels among countries. The Streamlit-based application makes it easier for users to understand global development patterns interactively. The results of this study are expected to serve as a basis for developing effective and equitable development policies.

Keywords: K-Means, Streamlit, GDP, Health, Clustering

PENDAHULUAN

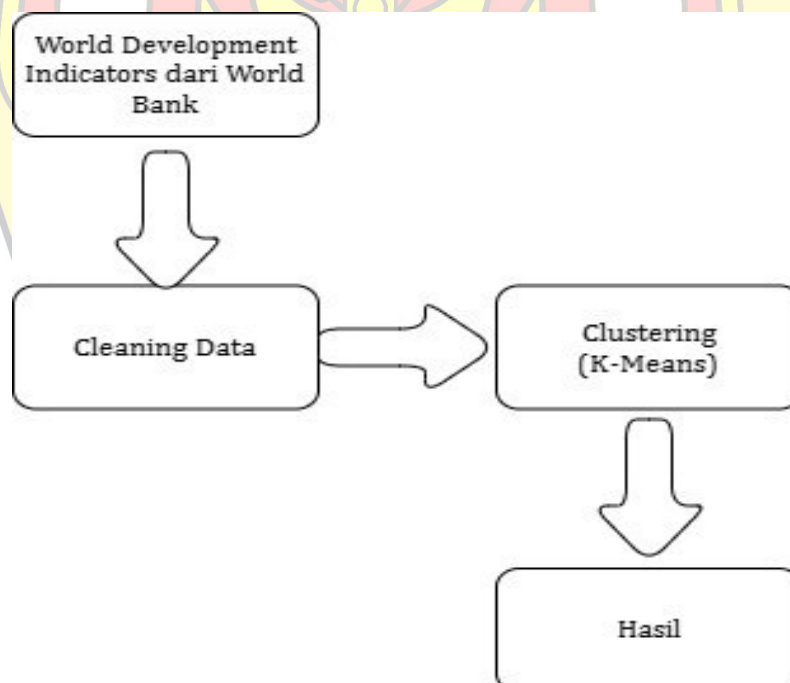
Pembangunan suatu negara ditentukan tidak hanya oleh kapasitas ekonominya, tetapi juga oleh kualitas hidup masyarakatnya, yang tercermin pada indikator kesehatannya. Penilaian kemajuan ekonomi suatu negara diukur dari PDB (Produk Domestik Bruto), dimana PDB berfungsi sebagai indikator utama untuk menentukan tingkat kesejahteraan ekonomi dan daya beli masyarakat (Negara et al., 2020). Namun, PDB saja tidak cukup untuk mencerminkan kesejahteraan masyarakat secara keseluruhan. Oleh karena itu, layanan kesehatan masyarakat juga penting untuk analisis yang komprehensif dalam suatu perkembangan negara. Penelitian yang mengintegrasikan faktor ekonomi dan kesehatan dalam satu analisis masih jarang dilakukan, meskipun keduanya saling terkait dan sangat mempengaruhi kebijakan publik (Romantika Marpaung et al., 2025). Banyak contoh di berbagai negara berkembang menunjukkan bahwa meskipun pertumbuhan ekonomi tinggi, hal itu tidak selalu berdampak langsung pada peningkatan kesejahteraan masyarakatnya (Nurhuda et al., 2025).

Ini menunjukkan bahwa model pembangunan internasional tidak selalu bersifat linear dan memerlukan pendekatan analitis yang lebih komprehensif. Stabilitas ekonomi adalah salah satu tujuan yang dikejar oleh setiap pemerintah di setiap negara. Indikator ini bersifat mendasar, yang mencerminkan kesehatan ekonomi suatu negara, dan berfungsi sebagai alat untuk menilai keberhasilan atau kegagalan pembangunan di suatu negara (Dian Merini, 2013). Untuk melihat apakah negara tersebut sudah berkembang penelitian ini menggunakan metode pengklasifikasian K-Means yang dapat mengkategorikan negara-negara di seluruh dunia. K-Means adalah suatu algoritma yang digunakan untuk mengelompokkan data berdasarkan kesamaan atributnya, metode pengelompokan dengan menggunakan algoritma K-Means diterapkan untuk meningkatkan segmentasi pola kekerasan terhadap perempuan dan anak-anak (Fithryani et al., 2025). Langkah pendekatan ini diharapkan dapat memberikan pengelompokan negara yang lebih akurat mengenai posisi dan kebutuhan dalam upaya mencapai pembangunan yang

berkelanjutan dan adil. Dengan menggunakan kombinasi Analisis Komponen Utama, K-Means, Streamlit, dan analisis kluster, akan mengidentifikasi pola kluster setiap negara dengan karakteristik yang serupa dan mengungkap fitur tersembunyi dalam data yang digunakan dalam penelitian ini. Streamlit adalah pustaka Python sumber terbuka yang memudahkan pembuatan dan berbagi aplikasi web kustom yang menarik untuk pembelajaran mesin dan ilmu data. Dengan Streamlit, dasbor interaktif, grafik, dan peta dapat dibuat dengan cepat dan mudah untuk memahami dan membagikan data (Darmawan, 2024).

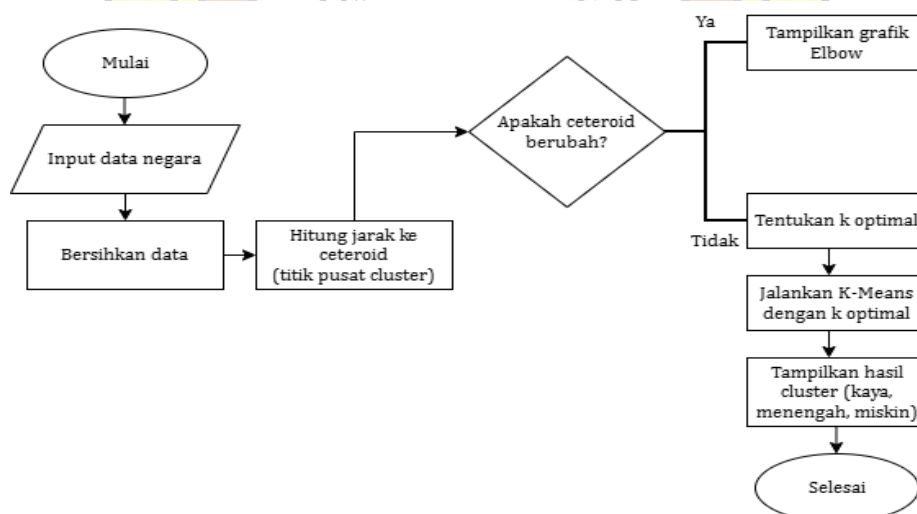
METODE PENELITIAN

Tahap persiapan data merupakan langkah utama dalam penelitian ini, dilakukan untuk memastikan kualitas dan kesesuaian data sebelum diproses. Proses ini sangat penting karena data seluruh dunia seringkali tidak lengkap, tidak konsisten, dan memiliki skala yang berbeda. Langkah-langkah pada penelitian yang akan dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian. (Hendrastuty, 2024).

Tahapan persiapan *World Development Indicators* Bank Dunia yang digunakan dalam penelitian ini adalah data sekunder. Data diperoleh dari Kaggle, (<https://www.kaggle.com/datasets/nicolasgonzalezmunoz/world-bank-world-development-indicators>) sebuah platform komunitas untuk data science yang menyediakan berbagai set data publik. Selain itu, studi ini menggunakan analisis kluster dan pendekatan kuantitatif. Seleksi Fitur Setelah dataset tersedia, tidak semua kolom fitur digunakan dalam analisis karena akan melakukan pembersihan data (*Data Cleaning*). Langkah ini mencakup pemeriksaan data untuk mengidentifikasi dan memperbaiki masalah yang sering terjadi. Proses yang dilakukan meliputi penanganan nilai yang hilang (*missing values*), baik dengan menghapus negara pada baris dengan data yang tidak lengkap maupun dengan menggunakan metode imputasi untuk melengkapi nilai yang hilang jika memungkinkan. Hal ini dilakukan untuk memastikan bahwa tidak ada data yang hilang yang akan mempengaruhi perhitungan algoritma klustering. Algoritma K-Means bekerja dengan mengukur negara berdasarkan PDB, harapan hidup, angka kelahiran, persentase akses listrik, dan pengeluaran kesehatan pemerintah, pendekatan klasifikasi ini akan mengungkap pola pada setiap pengelompokan negara. Algoritma ini sangat sensitif terhadap skala atau rentang nilai dari fitur yang digunakan.



Gambar 2. Diagram Penelitian.

Langkah berikutnya dari penelitian ini adalah memeriksa hubungan antara proses, pengambilan keputusan, dan aliran data dari awal hingga akhir agar lebih akurat. Pada penelitian ini menggunakan diagram alir (*flowcart*) seperti pada Gambar 2, yang merupakan tahap perancangan sistem, untuk membantu menganalisis kebutuhan logis sebelum menerapkan pola. Diagram alir adalah diagram yang digunakan untuk merepresentasikan alur logis atau langkah-langkah proses dari suatu sistem secara sistematis dan terstruktur. Dalam penelitian ini, diagram alir digunakan untuk menjelaskan proses algoritma K-Means dalam mengelompokkan negara berdasarkan tingkat kekayaannya (kaya, menengah, miskin), sehingga memberikan pemahaman langkah-langkah perhitungan yang dilakukan secara terstruktur dan visual.

K-Means membagi data menjadi k cluster dengan meminimalkan SSE (Sum of Squared Errors), dengan rumus:

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} ||x - \mu_i||^2$$

Keterangan:

k = jumlah cluster

C_i = kumpulan data pada cluster ke- i

μ_i = centroid (titik pusat) dari cluster ke- i

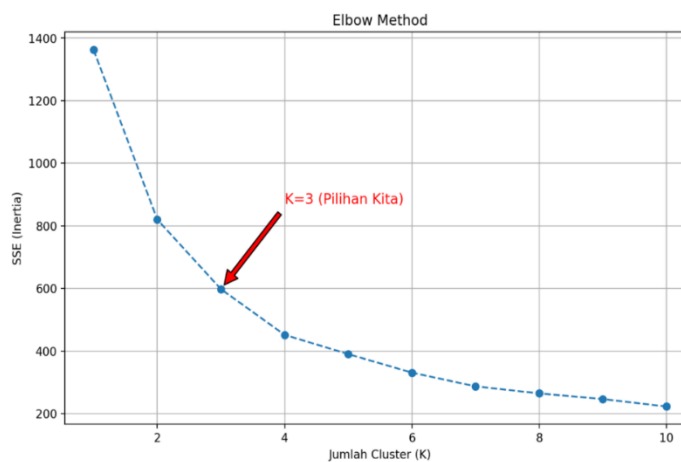
$||x - \mu_i||^2$ = jarak kuadrat antara data x dan centroid-nya

Sehingga ditemukanlah variabel yang akan digunakan dalam mengelompokkan suatu negara pada tabel berikut :

Tabel 1. Variabel

Variabel X	Variabel Y	Keterangan
GDP_current_US (PDB dalam Dolar AS)		Total nilai semua barang dan jasa yang diproduksi oleh suatu negara dalam setahun. Ini adalah ukuran utama untuk menentukan seberapa "kaya" atau besar ekonomi suatu negara.
	life_expectancy_at_birth (Angka Harapan Hidup)	Rata-rata jumlah tahun yang diharapkan untuk hidup oleh seorang bayi yang baru lahir. Ini mencerminkan kualitas sistem kesehatan, gizi, dan kondisi hidup secara umum.
	birth_rate (Angka Kelahiran)	Jumlah kelahiran hidup per 1.000 orang dalam satu tahun. Ini menunjukkan seberapa cepat populasi bertambah.
	death_rate (Angka Kematian)	Jumlah kematian per 1.000 orang dalam satu tahun. Ini dipengaruhi oleh kualitas layanan kesehatan, penyakit, dan usia populasi.
	access_to_electricity% (Akses ke Listrik %)	penduduk yang memiliki akses ke listrik. Ini adalah ukuran dasar infrastruktur dan kualitas hidup.
	government_health_expenditure% (Pengeluaran Kesehatan Pemerintah)	Persentase dari total PDB (GDP) suatu negara yang dibelanjakan oleh pemerintah untuk sektor kesehatan. Ini menunjukkan prioritas pemerintah pada kesehatan.

Instrumen variabel penelitian ini akan mengumpulkan data komprehensif yang mengukur kekuatan ekonomi dan kesehatan setiap negara. Setelah proses ini tahap pengelompokan dan standarisasi dari data yang digunakan akan dianalisis lebih lanjut.

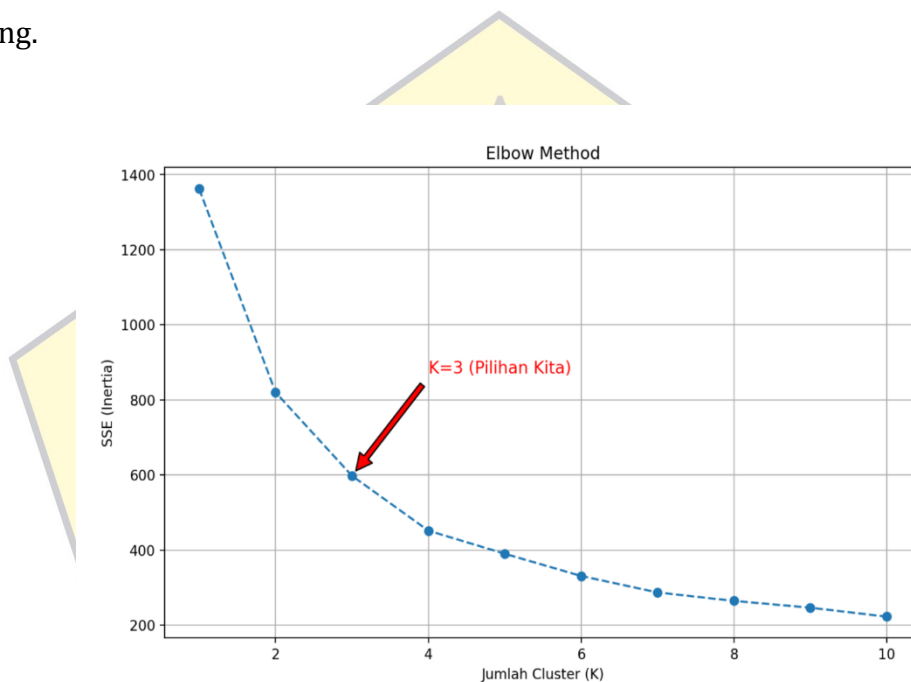


Gambar 3. Penerapan Elbow

Langkah berikutnya adalah menggunakan metode Elbow untuk menentukan jumlah cluster data. Metode ini digunakan untuk mendapatkan jumlah cluster yang optimal untuk sekumpulan titik data dengan cara yang sederhana dan mudah diterapkan. Dengan menggunakan algoritma clustering K-Means, metode Elbow dapat menggambarkan variasi dalam data yang dijelaskan oleh jumlah cluster dan memilih titik siku untuk menentukan jumlah cluster (Hartanti et al., 2024). Metode ini didasarkan pada gagasan bahwa seseorang harus memilih jumlah cluster sedemikian rupa sehingga menambahkan cluster lain tidak menghasilkan pemodelan data yang secara signifikan lebih baik (Bholowalia & Kumar, 2014). (This method is based on the idea that one should choose a number of clusters such that adding additional clusters does not significantly improve data modeling). Metode ini didasarkan pada prinsip bahwa seseorang harus memilih jumlah klaster tertentu sehingga penambahan klaster lain tidak memberikan pemodelan data yang jauh lebih baik (Bholowalia & Kumar, 2014). Dari diagram SSE metode Elbow pada Gambar 3 terlihat bahwa pada $k=3$ terjadi penurunan terbesar dibandingkan dengan nilai k lainnya. Jumlah klaster selain $k=3$ menunjukkan penurunan yang stabil. Oleh karena itu dapat disimpulkan bahwa jumlah klaster optimal yang terbentuk adalah 3.

HASIL DAN PEMBAHASAN

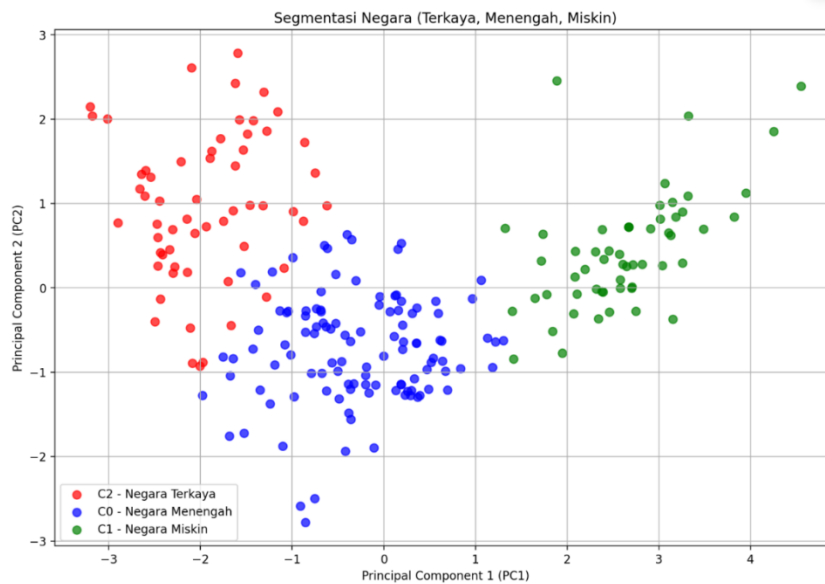
Pada diagram elbow yang digunakan untuk menentukan jumlah cluster (k) optimal dalam algoritma K-Means ketika negara dikelompokkan berdasarkan tingkat kesejahteraannya, sumbu horizontal (x) menunjukkan jumlah cluster k , sedangkan sumbu vertikal (y) menunjukkan nilai SSE (jumlah kuadrat selisih) atau inersia, yang mengukur total jarak kuadrat dari titik data ke pusat cluster masing-masing.



Gambar 4. Hasil Elbow

Gambar menunjukkan bahwa nilai SSE menurun secara signifikan ketika k meningkat dari 1 menjadi 3, yang menunjukkan bahwa pembagian data ke dalam lebih banyak cluster meningkatkan homogenitas di setiap cluster dan mengurangi jarak data ke pusat cluster. Namun, setelah $k=3$, penurunan nilai SSE secara bertahap melambat dan cenderung mendatar, yang dalam grafik disebut sebagai 'titik siku'. Titik ini menunjukkan bahwa peningkatan jumlah cluster lebih dari $k=3$ tidak memberikan perbaikan substansial dalam pengelompokan data yang efektif. Oleh karena itu, $k=3$ dipilih sebagai jumlah cluster yang optimal. Dengan demikian, data negara dibagi menjadi tiga cluster yang paling representatif, kemungkinan

sesuai dengan kategori kaya, menengah, dan miskin. Pemilihan jumlah cluster penting untuk memastikan bahwa model K-Means memberikan pengelompokan yang akurat dan bermakna sekaligus menghindari overfitting akibat terlalu banyak cluster.



Gambar 5. Hasil Cluster PCA

Hasil analisis menunjukkan bahwa klasifikasi ke dalam kategori C2 (Negara Berpenghasilan Tinggi), C1 (Negara Berpenghasilan Menengah), dan C0 (Negara Berpenghasilan Rendah) meningkatkan pemahaman secara keseluruhan (Carita et al., 2025). PCA adalah teknik analisis statistik yang sering dipakai untuk mengurangi dimensi data. Pendekatan ini muncul pertama kali di awal abad 20 dan terus mengalami perkembangan seiring dengan bertambahnya kompleksitas data yang dihadapi oleh ilmuwan dan praktisi di berbagai ilmu (Komprensif et al., 2025). Dalam penelitian ini menggunakan 6 variabel untuk mengelompokkan negara antara lain PDB, angka harapan hidup, angka kelahiran, angka kematian, akses listrik, dan pengeluaran kesehatan. Ini berarti variabelnya ada dalam ruang 6 dimensi. Mustahil untuk menggambar atau melihat data 6 dimensi pada layar komputer yang hanya punya sumbu X dan Y, Sehingga PCA adalah solusinya, disini kita masukkan 6 variabel ke dalam PCA, dan PCA akan menyatukan 6 variabel

tersebut menjadi 2 fitur buatan baru yang disebut Principal Component 1 (PC1) dan Principal Component 2 (PC2). (Priyanto & Tahyudin, 2024). Principal Component 1 (PC1) adalah indeks komposit atau ringkasan yang paling penting yang dihasilkan oleh metode PCA. Sebaliknya, PC1 adalah sebuah variabel baru yang dibentuk dari kombinasi linier semua variabel asli. Tujuan utama PC1 adalah untuk menangkap dan menjelaskan bagian terbesar dari perbedaan yang ada di seluruh dataset. Dalam studi klasterisasi, PC1 sering diartikan sebagai indeks pembangunan utama, karena ini adalah sumbu tunggal yang paling baik dalam memisahkan negara-negara berdasarkan perbedaan gabungan dari semua indikator yang digunakan (Bhahari, 2024). Sedangkan Principal Component 2 (PC2) adalah ringkasan terpenting kedua dari data. Setelah PC1 mengambil perbedaan terbesar dalam data, tugas PC2 adalah menangkap sisa varian terbesar berikutnya yang tidak dijelaskan oleh PC1. Secara konsep Independen dari PC1 adalah Informasi yang dibawa oleh PC2 sepenuhnya tegak lurus dengan informasi di PC1 (*E-Issn : 2988-1986*, 2025). Hasil PCA nya mengacu pada grafik titik-titik yang dimiliki. Hasil PCA adalah sumbu X dan sumbu Y dari grafik tersebut. Sumbu X = Principal Component 1 (PC1) Sumbu Y = Principal Component 2 (PC2).

PCA juga bisa memvisualisasikan hasil K-Means. Grafik PCA tersebut membuktikan bahwa 3 cluster memang berbeda. Titik-titik di Grafik menunjukkan sebuah negara, Titiknya juga memiliki Warna (merah=Terkaya, biru=Menengah, hijau=miskin) yang menunjukkan cluster hasil K-Means (C2-Terkaya, C0-Menengah, C1-Termiskin). Pada Gambar, Sumbu X (PC1) Lihat bagaimana cluster C2, C0, dan C1 terpisah jelas dari kiri ke kanan di sepanjang sumbu X. Ini berarti PC1 kemungkinan besar mewakili Indeks Pembangunan atau Kemakmuran. Negara dengan skor PC1 tinggi di kiri adalah negara kaya, dan negara dengan skor PC1 rendah di kanan adalah negara miskin. Sementara pada Sumbu Y (PC2) Sumbu ini menunjukkan perbedaan lain. Misalnya, dua negara menengah atau biru mungkin punya skor PC1 yang mirip yaitu sama-sama ditengah, tapi skor PC2-nya bisa beda satu di atas dan satu di bawah. Ini berarti mereka menengah dengan cara yang berbeda, mungkin

satu unggul di kesehatan tapi kurang di listrik, dan sebaliknya. Untuk menemukan pembagian negara beserta penghasilannya penelitian ini menggunakan rumus

$$z = \frac{(x - \mu)}{\sigma}$$

Keterangan:

x = nilai data asli,

μ = rata-rata (mean)

σ = standar deviasi

Untuk negara-negara yang telah diklasterkan peneliti mengambil 10 negara dari negara-negara yang telah diklasterkan berikut adalah tabel negara-negara tersebut:

Tabel 1. Negara Terkaya

Negara	PDB (GDP_current_US)
United States	21.060.470.000.000 (21 Triliun)
Japan	5.055.587.000.000 (5 Triliun)
Germany	3.887.727.000.000 (3,8 Triliun)
France	2.647.419.000.000 (2,6 Triliun)
Italy	1.897.462.000.000 (1,8 Triliun)
Canada	1.655.685.000.000 (1,6 Triliun)
Argentina	385.740.500.000 (385 Miliar)
Portugal	229.031.900.000 (229 Miliar)
New Zealand	212.569.800.000 (212 Miliar)
Albania	15.162.730.000 (15 Miliar)

Tabel 2. Negara Menengah

Negara	PDB (GDP_current_US)
China	14.687.740.000.000 (14,6 Triliun)
India	2.671.595.000.000 (2,6 Triliun)
Brazil	1.476.107.000.000 (1,4 Triliun)
Indonesia	1.059.055.000.000 (1,0 Triliun)
Malaysia	337.456.200.000 (337 Miliar)
Iraq	180.924.100.000 (180 Miliar)
Kuwait	105.948.800.000 (105 Miliar)
Bahrain	34.621.810.000 (34 Miliar)
Jamaica	13.812.420.000 (13,8 Miliar)
Brunei Darussalam	12.005.800.000 (12 Miliar)

Tabel 3. Negara miskin

Negara	PDB (GDP_current_US)
Kenya	100.657.500.000 (100 Miliar)
Ghana	70.043.100.000 (70 Miliar)
Papua New Guinea	23.848.450.000 (23 Miliar)
Zimbabwe	21.509.700.000 (21 Miliar)
Afghanistan	19.955.930.000 (19 Miliar)
Benin	15.686.740.000 (15 Miliar)
Guinea	14.177.840.000 (14 Miliar)
Malawi	12.056.110.000 (12 Miliar)
Congo, Rep.	11.468.610.000 (11 Miliar)
Burundi	2.649.680.000 (2,6 Miliar)

SIMPULAN

Studi ini bertujuan untuk mengelompokkan negara-negara di dunia berdasarkan indikator ekonomi dan kesehatan menggunakan algoritma K-Means yang diimplementasikan melalui aplikasi Streamlit. Analisis dengan menggunakan metode Elbow menunjukkan bahwa jumlah cluster yang optimal adalah tiga ($k = 3$). Ini berarti negara-negara dapat dikategorikan menjadi tiga kelompok utama: negara kaya, negara berpendapatan menengah, dan negara miskin. Teknik PCA digunakan untuk memvisualisasikan hasil pengelompokan ini dengan jelas, menyoroti perbedaan yang mencolok antara kelompok berdasarkan kombinasi variabel seperti PDB, harapan hidup, angka kelahiran, angka kematian, akses terhadap listrik, dan pengeluaran kesehatan. Aplikasi berbasis Streamlit ini juga membantu menyajikan hasil analisis secara interaktif, memungkinkan pengguna memahami posisi setiap negara dengan lebih baik. Secara keseluruhan, penelitian ini menunjukkan bahwa kombinasi indikator ekonomi dan kesehatan dapat memberikan gambaran yang lebih komprehensif tentang kesejahteraan suatu negara dan dapat menjadi dasar untuk mendukung proses pengambilan keputusan serta kebijakan pembangunan yang terarah.

UCAPAN TERIMA KASIH

Peneliti dengan penuh rasa syukur mengucapkan terima kasih yang sebesar-besarnya kepada Yayasan Universitas Bina Sarana Informatika atas segala dukungan, bantuan, dan fasilitas yang telah diberikan selama proses pengerjaan penelitian ini. Dukungan dari yayasan telah memberikan kontribusi yang sangat berarti bagi kelancaran penelitian, baik dalam bentuk bimbingan akademik maupun penyediaan sarana yang mendukung keberhasilan penelitian ini. Ucapan terima kasih juga peneliti sampaikan kepada rekan satu team yang sangat berpartisipasi atas seluruh kerjasama yang senantiasa menyediakan waktu, semangat, serta ilmu pengetahuan yang menjadi dasar penting dalam penyusunan penelitian ini.

DAFTAR PUSTAKA

- Carita, H. R., Hakim, H. K., Saputra, R. N. H., Fabiola, N. K. M., Albar, I. S., Rohman, V. A., & Nuraya, A. S. (2025). Dampak Perubahan Anggaran terhadap Output dan Pertumbuhan Ekonomi di Indonesia dalam Sisi Income and Spending. *JSE: Jurnal Sharia Economica*, 4(3), 158–174. <https://doi.org/10.46773/jse.v4i3.2148>
- Darmawan, D. (2024). Metodologi Penelitian Dalam Pengembangan Aplikasi Streamlit: Studi Kasus Visualisasi Data Gempa Bumi Di Indonesia. *Research Gate, August*. <https://doi.org/10.13140/RG.2.2.26320.32008>
- Dian Merini. (2013). Analisa Efisiensi Pengeluaran Pemerintah Sektor Publik di Kawasan Asia Tenggara: Aplikasi Data Envelopment Analysis. *Working Paper Universitas Brawijaya*, 1–31.
- Fithryani, N. M., Dikananda, A. R., Rohman, D., & Cirebon, K. (2025). *Algoritma K-Means Untuk*. 13(1), 997–1003.
- Negara, P., Indikator, B., Valentino, M., Sipahutar, Y., & Farhan, M. (2020). *PEMBANGUNAN GLOBAL MENGGUNAKAN METODE PCA DAN CLUSTERING K-MEANS TAHUN 2000-2020*.
- Nurhudaya, R. P., Bustomi, I., Danutirta, R., Ardeliana, L., Rudianto, B., & Alfian, Z. (2025). Analisis Perbandingan Hasil K-Means Clustering untuk Sektor Ekonomi dan Kesehatan. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(2), 4479–4487. <https://doi.org/10.31004/riggs.v4i2.1258>
- Romantika Marpaung, D., Gunawan, E., Rio Fa, F., & Christianto, A. (2025). Klasterisasi Negara Dunia Berdasarkan Data Sosioekonomi dan Demografi Tahun 2023 dengan PCA dan K-Means. *KOMPUTA : Jurnal Ilmiah Komputer Dan Informatika*, 14(1), 2715–7849. <https://doi.org/10.34010/komputa.v14i1>.
- Bholowalia, P., & Kumar, A. (2014). EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. *International Journal of Computer Applications*, 105(9), 975–8887.
- Hartanti, N. T., Seniwati, E., Pramitasari, R., Studi, P., Informasi, S., Amikom, U., Studi, P., Komputer, T., & Amikom, U. (2024). Metode Elbow K-Means dalam Implementasi Data Mining pada Pemetaan Penyebaran Guru SMK. *Teknika*, 18(2), 501–512.
- Hendrastuty, N. (2024). Pengertian Analisis Data. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (jima-Ilkom)*, 3(1), 46–56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>

- Rianti, R., Andarsyah, R., & Awangga, R. M. (2024). Penerapan PCA dan Algoritma Clustering untuk Analisis Mutu Perguruan Tinggi di LLDIKTI Wilayah IV. *Nuansa Informatika*, 18(2), 67–77. <https://doi.org/10.25134/ilkom.v18i2.211>
- Sarimole, F. M., & Hakim, L. (2024). Klasifikasi Barang Menggunakan Metode Clustering K-Means Dalam Penentuan Prediksi Stok Barang. *Jurnal Sains Dan Teknologi*, 5(3), 846–854. <https://doi.org/10.55338/saintek.v5i3.2709>
- Turnip, M., Setiawan, H., Sitanggang, D., & Aisyah, S. (n.d.). *Implementasi Metode K-Means Clustering Untuk*. 9, 80–89.
- Bhahari, R. H. (2024). *Clustering Analysis of Socio-Economic Districts / Cities In East Java Province Using PCA And Hierarchical Clustering Methods*. 8(4), 2242–2251. *E-issn : 2988-1986*. (2025). 8(7).
- Komprehensif, T., Principal, P., & Analysis, C. (2025). *Jurnal EurekaMatika*. 13(1), 25–34.
- Priyanto, E., & Tahyudin, I. (2024). *Analisis Relevansi Kompetensi Alumni dengan Pekerjaan di Pendidikan Tinggi Menggunakan Pendekatan PCA dan Clustering Magister of Computer Science , Universitas Amikom Purwokerto , Jawa Tengah , Indonesia The Application of PCA for Dimensionality Reduction in K-Means Clustering for Analyzing the Relevance of Education and Employment in the Higher Education Sector*. 4(12), 785–793.