

# APLIKASI SEGMENTASI NEGARA BERDASARKAN KONTRIBUSI DAN DAMPAK PEMANASAN GLOBAL MENGGUNAKAN ALGORITMA K-MEANS

Muhammad Ridho Pratama<sup>1</sup>, Jatmalika Abi Kurniawan<sup>2</sup>, Muhammad Rizal<sup>3</sup>, Wahyu Rivaldi<sup>4</sup>, Siti Nurdiani<sup>5</sup>

<sup>1,2,3,4,5</sup>Universitas Bina Sarana Informatika Kampus Pontianak; Jl. Abdul Rahman Saleh No.18, Bangka Belitung Laut, Kec. Pontianak Tenggara, Kota Pontianak, Kalimantan Barat 78124, Telp. (0561) 583924

## Keywords:

Global Warming, K-Means Clustering, Streamlit, Elbow, Segment Counties.

## Correspondent Email:

15230386@bsi.ac.id

**Abstrak:** Pemanasan global menuntun pada pendekatan berbasis data untuk menyusun kontribusi dan dampak bagi setiap negara secara akurat. Penelitian ini menggunakan algoritma K-Means Klustering untuk mensegmentasikan negara yang diterapkan melalui platform Streamlit. Dataset perubahan iklim diolah melalui beberapa tahapan preprocessing, termasuk normalisasi data menggunakan StandardScaler guna menghindari ketimpangan atribut tertentu. Berdasarkan analisis dengan metode Elbow, diperoleh tiga kluster yang terbaik ( $K = 3$ ) yang mewakili kelompok Resiko Tinggi, Mitigasi Aktif, dan Kontribusi Polusi Tinggi. Penelitian ini membuktikan bahwa K-Means secara efektif mampu untuk mengelompokkan data yang bersifat multidimensi. Selain itu, Streamlit memberikan fasilitas berupa visualisasi interaktif dan mendukung penyajian data secara dinamis. Penelitian ini diharapkan mampu memberikan perspektif yang komprehensif bagi para pemangku kebijakan dalam menyusun strategi mitigasi iklim yang lebih terarah dan efektif.



Copyright © [JITET](http://www.jitet.org) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

**Abstract:** Global warming leads to a data-driven approach for accurately compiling the contributions and impacts for each country. This research uses the K-Means Clustering algorithm to segment countries, implemented through the Streamlit platform. The climate change dataset is processed through several preprocessing stages, including data normalization using StandardScaler to prevent disparity in certain attributes. Based on analysis with the Elbow method, three optimal clusters were obtained ( $K = 3$ ), representing the groups: High Risk, Active Mitigation, and High Pollution Contribution. This study proves that K-Means can effectively cluster multidimensional data. Furthermore, Streamlit provides interactive visualization facilities and supports dynamic data presentation. This research is expected to provide a comprehensive perspective for policymakers in formulating more focused and effective climate mitigation strategies.

## 1. PENDAHULUAN

Pemanasan global telah menjadi isu yang mendesak dan memerlukan perhatian serius dari seluruh negara di dunia. Dampak dari fenomena ini bersifat multidimensional dan

tidak hanya mempengaruhi aspek fisik lingkungan, tetapi juga sistem sosial, ekonomi, dan kesehatan global [1]. Pemanasan global menimbulkan dampak yang luas dan kompleks, mulai dari curah hujan yang tidak menentu,

kekeringan, kegagalan panen, kerusakan ekosistem laut, hingga kepunahan spesies. Fenomena ini juga berkontribusi terhadap penipisan lapisan ozon serta mencairnya es di kutub yang memicu kenaikan permukaan laut, sehingga mengancam negara kepulauan dan negara berkembang. Selain itu, perubahan iklim akibat pemanasan global memunculkan bencana hidrometeorologi seperti banjir, kekeringan, tanah longsor, dan angin puting beliung [2].

Guna memahami dan mengatasi permasalahan tersebut, diperlukan pendekatan secara berbasis data yang kuat. Di sinilah machine learning digunakan sebagai salah satu cabang ilmu dari *artificial intelligence* yang menggunakan beragam statistik, teknik portabilitas, dan optimasi variabel yang berbeda atau menemukan ketidakseimbangan data. Machine learning mencakup perancangan dan pengembangan algoritma agar komputer dapat mengembangkan perilaku berdasarkan data yang sudah diberikan [3].

Untuk memahami dan mengelompokkan negara-negara berdasarkan kontribusi dan dampak pemanasan global, diperlukan pendekatan analisis data yang tepat. Salah satu metode yang digunakan adalah K-Means Klustering, yaitu metode pengelompokan data non-hierarki yang membagi data ke dalam satu atau lebih klaster berdasarkan kemiripan karakteristik. Data yang memiliki karakteristik serupa akan dikelompokkan dalam satu klaster, sedangkan data yang berbeda akan masuk ke klaster lain. Hasil pengelompokan ini dapat menjadi informasi yang bermanfaat dalam proses pengambilan keputusan, terutama dalam merumuskan kebijakan lingkungan dan strategi mitigasi global [4].

Untuk mempermudah pengolahan dan penyajian data, diperlukan Streamlit sebagai framework *open-source* berbasis Python yang digunakan untuk pengembangan aplikasi web untuk keperluan data science dan machine learning dengan cepat dan efisien. Streamlit dapat digunakan dalam skenario seperti pembuatan prototipe sistem berbasis kecerdasan buatan, presentasi hasil analisis data, hingga pembuatan dashboard yang bersifat dinamis [5].

Sejumlah penelitian telah membuktikan efektivitas metode K-Means Klustering dalam berbagai bidang analisis data. Salah satu

penelitian berjudul “Implementasi Metode K-Means Klustering untuk Pengelompokan Generasi Milenial Berdasarkan Perilaku Hemat Energi” menunjukkan bahwa algoritma ini mampu mengidentifikasi pola kesadaran, sikap, dan tindakan generasi milenial dalam menerapkan kebiasaan hemat energi. Temuan tersebut dapat dimanfaatkan untuk merumuskan strategi peningkatan kesadaran energi yang lebih efektif. Penelitian ini juga menegaskan keunggulan K-Means dalam hal kemudahan implementasi, kecepatan pemrosesan, serta efisiensi dalam menangani dataset berukuran besar [6].

Oleh Karena itu, penelitian ini bertujuan untuk mengembangkan aplikasi segmentasi negara berdasarkan kontribusi dan dampak pemanasan global menggunakan algoritma K-Means Klustering yang diimplementasikan melalui platform Streamlit. Melalui integrasi metode K-Means dengan platform Streamlit, penelitian ini diharapkan dapat membantu visualisasi serta analisis data lingkungan secara interaktif, sehingga mendukung proses pengambilan keputusan dan penyusunan kebijakan mitigasi yang lebih tepat sasaran.

## 2. TINJAUAN PUSTAKA

### 2.1. Pemanasan Global

Pemanasan global adalah meningkatnya suhu rata-rata permukaan bumi yang diakibatkan oleh aktivitas manusia, seperti pembakaran bahan bakar fosil yang melepaskan karbon dioksida, alih fungsi lahan, serta pelepasan gas-gas lain yang dikenal sebagai gas rumah kaca. Akumulasi gas-gas ini menyebabkan efek rumah kaca berlebih di atmosfer bumi, yang diyakini menjadi salah satu penyebab utama terjadinya perubahan iklim secara ekstrem [7].

### 2.2. K-Means Klustering

K-Means Klustering merupakan metode pengelompokan data non-hierarki yang membagi dataset ke dalam K klaster berdasarkan kemiripan karakteristik antar data [4]. K-means menggunakan proses secara berulang-ulang untuk mendapatkan basis data klaster. Dibutuhkan jumlah custer awal yang diinginkan sebagai masukan dan menghasilkan titik *centroid* akhir sebagai output. Metode K-means akan memilih pola k sebagai titik awal *centroid* secara acak atau random. Jumlah iterasi untuk mencapai klaster *centroid*

akan dipengaruhi oleh calon kluster *centroid* awal secara random. Sehingga didapat cara dalam pengembangan algoritma dengan menentukan *centroid* kluster yang dilihat dari kepadatan data awal yang tinggi agar mendapatkan kinerja yang lebih tinggi [8].

### 2.3. Preprocessing Data

Preprocessing data merupakan tahapan krusial dalam proses data mining yang bertujuan meningkatkan kualitas data dari segi kelengkapan, konsistensi, dan ketepatan waktu [9]. Tahapan ini meliputi penanganan nilai hilang (missing values), penghapusan data duplikat, deteksi dan penanganan *outlier*, serta transformasi data. Dalam konteks klustering, preprocessing yang tepat sangat menentukan kualitas hasil pengelompokan.

### 2.4. Seleksi fitur

Seleksi fitur adalah proses memilih subset fitur yang paling relevan dan informatif untuk digunakan dalam pemodelan [10]. Tujuan seleksi fitur adalah mengurangi dimensionalitas data, meningkatkan akurasi model, mengurangi waktu komputasi, serta menghindari overfitting. Dalam penelitian ini, fitur-fitur yang dipilih adalah yang paling mencerminkan kontribusi dan dampak pemanasan global, yaitu emisi CO<sub>2</sub> per kapita, penggunaan energi terbarukan, dan suhu rata-rata.

### 2.5. Silhouette Score

*Silhouette Score* berfungsi sebagai metrik evaluasi untuk menilai sejauh mana suatu objek terasosiasi secara tepat dengan kluster tempatnya berada, sekaligus merefleksikan koherensi dan kekompakan kluster tersebut. Nilai *Silhouette* berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi mengindikasikan kualitas pengelompokan yang lebih optimal. *Silhouette Score* memiliki rentang nilai antara -1 hingga 1, dengan interpretasi sebagai berikut [11]:

1. Nilai mendekati 1 menunjukkan bahwa suatu observasi terkelompok secara optimal dalam kluster yang sesuai.
2. Nilai mendekati 0 mengindikasikan bahwa observasi berada di batas antara dua kluster
3. Sedangkan nilai negatif merefleksikan kemungkinan kesalahan penempatan observasi dalam kluster yang tidak tepat.

Kalkulasi *Silhouette Score* ditunjukkan pada persamaan :

$$s(i) = \frac{b(i)-a(i)}{\max \{a(i),b(i)\}} \quad (1)$$

### 2.6. Metode Elbow

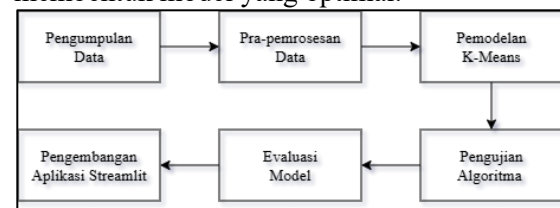
Metode Elbow merupakan salah satu teknik evaluasi yang digunakan untuk menentukan nilai *k* optimum dalam proses klusterisasi. Penentuan nilai tersebut dilakukan dengan membandingkan hasil perhitungan *Sum of Square Error (SSE)* pada berbagai iterasi, yang kemudian disajikan dalam bentuk grafik. Titik optimum dicapai ketika grafik menunjukkan adanya “patahan” atau perubahan signifikan pada kurva, yang dikenal sebagai kondisi Elbow [12].

### 2.7. Streamlit

Streamlit adalah framework *open-source* berbasis Python yang dirancang khusus untuk mengembangkan aplikasi web interaktif dalam bidang data science dan machine learning . Framework ini memungkinkan peneliti dan praktisi untuk membuat aplikasi web dengan cepat tanpa memerlukan pengetahuan mendalam tentang HTML, CSS, atau JavaScript. Streamlit dapat digunakan dalam skenario seperti pembuatan prototipe sistem berbasis kecerdasan buatan, presentasi hasil analisis data, hingga pembuatan dashboard yang bersifat dinamis [5].

## 3. METODE PENELITIAN

Pada tahap ini dimulai dengan pengumpulan dataset hingga pengembangan aplikasi Streamlit, Berikut tahap spesifik agar membentuk model yang optimal.



Gambar 1. Diagram Tahapan Penelitian

### 3.1. Pengumpulan Data

Dataset yang digunakan berasal dari platform Kaggle dengan judul *Climate Change Dataset*. Dataset ini berisi berbagai indikator lingkungan dari berbagai negara, seperti suhu rata-rata tahunan, emisi karbondioksida per kapita, kenaikan permukaan laut, curah hujan, proporsi energi terbarukan dan luas area hutan. Dataset ini dipilih karena relevan terhadap objek penelitian saat ini.

Dataset ini terdiri dari 1000 entri dengan sepuluh atribut utama, seperti *Year*, *Country*, *Average Temperature (°C)*, *CO2 Emissions (Tons/Capita)*, *Sea Level Rise (mm)*, *Rainfall (mm)*, *Population*, *Renewable Energy (%)*, *Extreme Weather Events*, dan *Forest Area (%)*. Atribut-atribut tersebut digunakan untuk mengelompokkan negara berdasarkan kontribusi dan dampaknya terhadap pemanasan global.

### 3.2. Preprocessing Data

Tahap preprocessing data merupakan Langkah penting dalam proses data mining yang memiliki tujuan untuk membuat kualitas data menjadi baik. meliputi kelengkapan, konsistensi, ketepatan waktu dan meningkatkan hasil akurasi. Pada tahap preprocessing dilakukan konversi data dengan mengeleminasi data yang hilang dan standarisasi data dengan tujuan menyamakan rentang nilai antar fitur, guna mengoptimalkan performa klasifikasi [9].

#### 1. Penanganan Missing Value

Langkah awal adalah melakukan pengecekan terhadap dataset untuk mendeteksi adanya nilai yang tidak tersedia atau kosong. Dalam penelitian ini, baris data yang mengandung nilai kosong pada kolom *CO2 Emissions (Tons/Capita)* dan *Renewable Energy (%)* dihilangkan dari dataset. Keputusan ini diambil berdasarkan pertimbangan bahwa kedua kolom tersebut merupakan variabel utama dalam proses pengelompokan. Oleh karena itu, pendekatan ini diterapkan untuk memastikan konsistensi dan meningkatkan akurasi hasil analisis, serta meminimalkan potensi bias yang dapat timbul akibat keberadaan data yang tidak lengkap.

#### 2. Pemilihan Variable

Dari sepuluh variable yang tersedia dalam dataset, ditetapkan dua variabel utama yang dianggap paling mencerminkan kontribusi dan pengaruh terhadap pemanasan global. Seleksi fitur digunakan untuk menentuakn fitur-fitur yang memiliki hubungan yang signifikan dengan variabel target dan mempertahankan fitur-fitur tersebut, sementara fitur yang tidak relevan akan dihilangkan guna menjaga kualitas dan efisiensi model.[10]

#### 3. Standarisasi Data

Proses standarisasi dilakukan dengan menggunakan *StandardScaler* yang mentransformasi nilai data menjadi distribusi dengan nilai 0 dan simpangan baku 1. Tahap

standarisasi ini memiliki peran penting untuk mencegah terjadinya dominasi variabel dengan rentang nilai yang lebih besar dalam menentukan hasil pengelompokan [13]. Formula *StandardScaler* dapat dilihat pada persamaan :

$$z = \frac{x - \mu}{\sigma} \quad (2)$$

### 3.3. Pemodelan K-Means

Dalam penelitian ini, pemilihan algoritma K-Means didasarkan pada tiga karakteristik utama yang dianggap krusial [14].

1. Algoritma ini memiliki struktur yang sederhana serta mudah diimplementasikan, sehingga mendukung proses analisis dan evaluasi model secara lebih efisien.
2. K-Means dikenal memiliki tingkat efisiensi komputasi yang tinggi, menjadikannya metode yang skalabel untuk pengolahan dataset berukuran sedang hingga besar dengan waktu yang relatif cepat.
3. Algoritma ini mampu bekerja secara optimal dalam mengelompokkan data dengan distribusi memusat atau berbentuk spherical, sehingga menghasilkan kluster yang padat sekaligus terpisah dengan baik.

Cara kerja algoritma ini adalah menentukan jumlah kelompok kluster (Nilai *K*). selanjutnya menginisialisasi titik pusat (*centroid*) sebanyak *K* buah secara acak atau berdasarkan metode tertentu. Setelah itu, dilakukan perhitungan jarak antara setiap data dengan seluruh *centroid* yang tersedia untuk menentukan kedekatan relative terhadap masing-masing kluster [15].

### 3.4. Pengujian Algoritma

Tahap pengujian dilakukan dengan menerapkan algoritma K-Means pada dataset penelitian. Proses dimulai dengan menentukan jumlah kluster (*K*) yang diinginkan, kemudian menetapkan centroid awal secara acak sesuai jumlah kluster tersebut. Selanjutnya, jarak antara setiap data dan *centroid* dihitung menggunakan rumus Euclidean Distance:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

Di mana *x* dan *y* merepresentasikan vektor atribut dari dua data yang dibandingkan. Berdasarkan hasil perhitungan jarak, data

dikelompokkan ke dalam kluster yang memiliki centroid terdekat. Posisi *centroid* kemudian diperbarui secara iteratif hingga tercapai kondisi konvergen, yaitu ketika perubahan posisi *centroid* tidak lagi signifikan.

Tahap ini menghasilkan pembentukan kluster sesuai parameter yang telah ditentukan, yang selanjutnya digunakan sebagai dasar dalam proses evaluasi model.

### 3.5. Evaluasi Model

Tahap evaluasi model klustering dilakukan untuk menilai kualitas hasil pengelompokan yang diperoleh melalui penerapan algoritma K-Means. Evaluasi ini bertujuan untuk mengukur seberapa efektif algoritma dalam membentuk kluster yang representatif terhadap struktur data yang dianalisis. Dalam penelitian ini, metrik yang digunakan untuk menilai performa pengelompokan adalah *Silhouette Score*. *Silhouette Score* berfungsi sebagai metrik evaluasi untuk menilai sejauh mana suatu objek terasosiasi secara tepat dengan kluster tempatnya berada, sekaligus merefleksikan koherensi dan kekompakan kluster tersebut. Nilai *Silhouette* berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi mengindikasikan kualitas pengelompokan yang lebih optimal [11].

### 3.6. Pengembangan Streamlit

Perancangan aplikasi segmentasi negara berdasarkan dampak pemanasan global menggunakan framework Streamlit yang berbasis web yang memungkinkan visualisasi dan interaksi dengan hasil klustering. Streamlit digunakan karena frameworknya yang open-source berbasis Python yang dirancang khusus untuk mengembangkan aplikasi web dalam bidang *data science* dan machine learning tanpa perlu memiliki pemahaman mendalam mengenai HTML, CSS dan JavaScript. Bagian ini akan memaparkan tahapan penelitian secara menyeluruh serta arsitektur system dari aplikasi yang dikembangkan.

## 4. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil eksperimen klusterisasi dengan menggunakan algoritma K-Means serta penerapannya pada aplikasi berbasis web Streamlit. Pembahasan meliputi uraian mengenai karakteristik data, analisis dalam menentukan jumlah kluster yang optimal, interpretasi terhadap ciri khas masing-masing kluster yang terbentuk, serta diskusi terkait

implikasi temuan dan efektivitas aplikasi yang dihasilkan.

### 4.1. Deskripsi Data dan Seleksi Fitur

Penelitian ini memanfaatkan *Climate Change Dataset* yang diperoleh dari repositori publik Kaggle. Dataset tersebut berisi 1.000 entri yang merekam beragam indikator lingkungan global. Untuk menjaga fokus analisis pada keterkaitan antara aktivitas manusia dan dampak lingkungan, dilakukan proses seleksi fitur (*feature selection*). Dari hasil seleksi, ditetapkan tiga variabel utama sebagai masukan algoritma karena memiliki tingkat relevansi yang tinggi terhadap isu pemanasan global.

Tabel. 1 Deskripsi Atribut (Fitur) Dataset

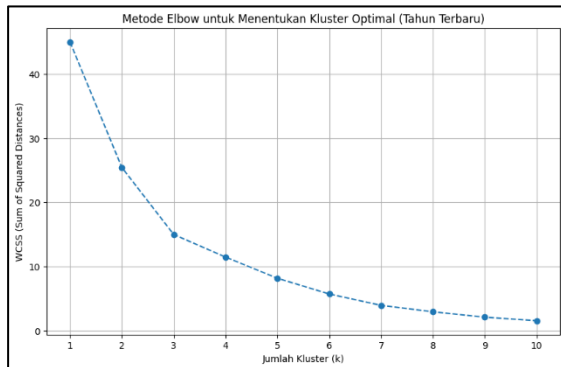
Nama Atribut	Keterangan	Contoh Nilai
CO <sup>2</sup> Emissions (Tons/Capita)	Indikator utama penyebab efek rumah kaca yang merepresentasikan kontribusi polusi suatu negara.	16.5
Renewable Energy (%)	Indikator upaya mitigasi yang menggambarkan transisi negara menuju energi bersih.	16.1
Avg Temperature (°C)	Indikator dampak fisik yang dirasakan akibat kenaikan suhu global.	28.5

Mengingat K-Means merupakan metode klusterisasi berbasis jarak (*Euclidean Distance*), seluruh variabel numerik terlebih dahulu dinormalisasi menggunakan *StandardScaler*. Proses normalisasi ini bertujuan untuk mencegah dominasi atribut dengan skala nilai besar, seperti populasi, sehingga perhitungan jarak antar data tetap proporsional dan menghasilkan kluster yang lebih akurat [16].

### 4.2. Penentuan Jumlah Kluster Optimal

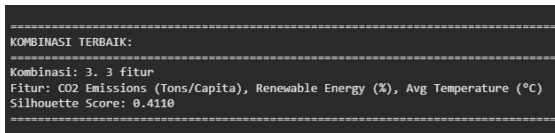
Langkah penting dalam algoritma K-Means adalah penentuan jumlah kluster (*K*) yang paling representatif. Dalam penelitian ini,

metode *Elbow* digunakan untuk menganalisis nilai Inertia atau *Within-Klaster Sum of Squares* (WCSS). Grafik pada Gambar 1 memperlihatkan perubahan nilai inersia terhadap variasi jumlah klaster.



Gambar 2. Metode *Elbow*

Seperti ditunjukkan pada Gambar 1, penurunan nilai WCSS yang signifikan terjadi hingga  $K = 3$ , kemudian kurva mulai melandai dan membentuk pola siku. Kondisi tersebut menunjukkan bahwa penambahan klaster lebih dari tiga tidak memberikan pengurangan variansi yang berarti. Validasi tambahan melalui *Silhouette Score* menghasilkan nilai sebesar 0,4110 untuk  $K = 3$ .



Gambar 3. *Silhouette Score*

Nilai positif tersebut menegaskan bahwa struktur klaster yang terbentuk memiliki tingkat pemisahan (*separation*) dan kepadatan (*cohesion*) yang memadai, sehingga model dapat dinyatakan valid untuk digunakan dalam analisis.

#### 4.3. Interpretasi Profil Klaster

Dengan menggunakan parameter  $K = 3$ , algoritma K-Means mengelompokkan data negara ke dalam tiga klaster yang memiliki karakteristik berbeda. *Centroid* yang terbentuk pada masing-masing klaster merepresentasikan profil khas dari setiap segmen, sebagaimana ditunjukkan dalam Tabel 2.

Tabel 2. Interpretasi Klaster

klaster	CO2 Emissions	Rene wable	Avg Temper
---------	---------------	------------	------------

	(Tons/Capit a)	Energ y (%)	ature (°C)
0	12.65	33.20	19.12
1	8.65	18.56	19.90
2	14.06	25.65	29.06

Berdasarkan nilai rata-rata yang ditampilkan pada Tabel 2, karakteristik masing-masing klaster dapat dijelaskan sebagai berikut:

1. Klaster 0 : Kelompok negara dalam klaster ini memiliki rata-rata emisi CO<sub>2</sub> sebesar 12,65 ton per kapita. Ciri utama yang menonjol adalah tingginya pemanfaatan energi terbarukan, dengan rata-rata mencapai 33,20%, jauh lebih tinggi dibandingkan klaster lainnya.
2. Klaster 1 : Klaster ini merepresentasikan negara-negara dengan jejak karbon paling kecil, ditunjukkan oleh rata-rata emisi hanya 8,65 ton per kapita serta suhu rata-rata yang relatif rendah, yaitu 19,90°C.
3. Klaster 2 : Klaster ini terdiri atas negara-negara dengan tingkat emisi tertinggi, yakni 14,06 ton per kapita, sekaligus suhu rata-rata tertinggi sebesar 29,06°C.

Analisis terhadap ketiga profil klaster menunjukkan pola yang konsisten, yakni tingginya tingkat emisi karbon berkorelasi dengan peningkatan suhu rata-rata. Sementara itu, proporsi penggunaan energi terbarukan menjadi variabel pembeda utama yang memisahkan karakteristik antar kelompok negara.

#### 4.4. Analisis Mendalam Profil Klaster

Hasil segmentasi tersebut memperlihatkan adanya pola keterkaitan yang signifikan antara emisi karbon, penggunaan energi, dan suhu rata-rata. Temuan pada Klaster 2 mendukung literatur sebelumnya yang menunjukkan bahwa tingginya emisi karbon memiliki korelasi positif dengan peningkatan suhu wilayah. Negara-negara dalam klaster ini menghadapi risiko ganda, yakni berperan sebagai penyumbang utama polusi sekaligus menjadi pihak yang paling terdampak oleh kenaikan suhu.

Sebaliknya, Klaster 0 memberikan perspektif strategis bahwa negara dengan tingkat emisi menengah mampu menekan dampak lingkungannya melalui penerapan



energi terbarukan secara intensif. Hal ini menegaskan efektivitas algoritma K-Means dalam mengidentifikasi pola kebijakan energi secara lebih mendalam. Tidak seperti metode statistik deskriptif konvensional, pendekatan klasterisasi ini berhasil menangkap dimensi yang lebih kompleks, di mana negara dinilai berdasarkan kombinasi antara perilaku emisi dan respons kebijakan energi terbarukan.

#### 4.5. Implementasi Aplikasi Streamlit

Untuk memperluas diseminasi hasil penelitian, model K-Means yang telah divalidasi diimplementasikan ke dalam aplikasi web berbasis framework Streamlit. Antarmuka aplikasi dirancang secara interaktif guna memudahkan pengguna dalam melakukan eksplorasi data.



Gambar 4. Antarmuka Input Aplikasi Streamlit

Sebagaimana ditunjukkan pada Gambar 4, aplikasi menyediakan fitur filter tahun yang memungkinkan analisis tren perubahan status klaster suatu negara dari waktu ke waktu. Visualisasi hasil ditampilkan secara real-time dalam bentuk scatter plot interaktif serta tabel data, sebagaimana diperlihatkan pada Gambar 3.



Gambar 5. Hasil Analisis Aplikasi Streamlit



Gambar 6. Tampilan Daftar Negara Per Klaster

Integrasi Streamlit dalam penelitian ini memberikan keunggulan yang signifikan dibandingkan metode penyajian data konvensional. Fitur *Search Country* pada dashboard memungkinkan pemangku kebijakan memantau posisi spesifik suatu negara secara cepat dan akurat. Selain itu, visualisasi berupa *scatter plot* (Gambar 3) memudahkan proses identifikasi terhadap *outlier* atau anomali data yang berpotensi terlewat jika hanya menggunakan tabel. Aplikasi ini berfungsi sebagai prototipe *Decision Support System* (DSS) yang efektif, dengan menyediakan akses data yang transparan dan interaktif untuk mendukung perumusan strategi mitigasi iklim yang lebih terarah.

## 5. KESIMPULAN

### A. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan, diperoleh beberapa kesimpulan sebagai berikut:

1. Algoritma K-Means Klastering terbukti mampu melakukan segmentasi negara berdasarkan indikator perubahan iklim multidimensi. Analisis menggunakan metode *Elbow* menunjukkan bahwa jumlah klaster optimal adalah  $K = 3$ .
2. Kualitas hasil pengelompokan telah diverifikasi melalui metrik *Silhouette Score* dengan nilai 0,4110. Nilai positif ini menandakan bahwa klaster yang terbentuk memiliki tingkat pemisahan (*separation*) dan kepadatan (*cohesion*) yang memadai, sehingga model dapat dinyatakan valid.
3. Penelitian berhasil mengidentifikasi tiga kelompok utama negara, yaitu: (1) Klaster Terdampak dan Rentan (ditandai dengan suhu tinggi dan kenaikan permukaan laut signifikan), (2) Klaster Mitigasi Aktif (ditandai dengan pemanfaatan energi

terbarukan yang tinggi), dan (3) Kluster Kontributor Utama (ditandai dengan emisi CO<sub>2</sub> yang tinggi).

4. Model K-Means diintegrasikan ke dalam aplikasi web berbasis Streamlit yang berfungsi sebagai *prototipe* sistem interaktif. Fitur seperti visualisasi scatter plot dan pencarian negara memungkinkan pemangku kepentingan memantau posisi serta kontribusi setiap negara terhadap pemanasan global secara *real-time*.

## B. Saran

Mengacu pada keterbatasan penelitian, terdapat beberapa rekomendasi untuk pengembangan di masa mendatang:

1. Disarankan untuk melakukan perbandingan dengan algoritma lain, seperti DBSCAN yang berbasis kepadatan maupun Gaussian Mixture Model (GMM), guna memvalidasi konsistensi hasil terutama dalam konteks data yang mengandung *outlier*.
2. Penelitian lanjutan sebaiknya memperkaya dataset dengan menambahkan variabel kualitatif, misalnya kebijakan iklim negara, yang dapat diolah menggunakan teknik *one-hot encoding* agar profil kluster lebih komprehensif.
3. Penerapan metrik validasi lain, seperti *Davies-Bouldin Index*, direkomendasikan untuk menilai kualitas pemisahan kluster secara lebih ketat.
4. Aplikasi berbasis Streamlit dapat ditingkatkan dengan fitur analitik lanjutan, seperti visualisasi tren *time-series* per kluster serta optimalisasi tampilan responsif untuk perangkat mobile, sehingga aksesibilitas data menjadi lebih luas.

## DAFTAR PUSTAKA

- [1] Jefri Kurniawan, Abdul Razak, Nurhasan Syah, Skunda Diliarosta, and Aulia Azhar, "Pemanasan Global: Faktor, Dampak dan Upaya Penanggulangan," *INSOLOGI J. Sains dan Teknol.*, vol. 3, no. 6, pp. 646–655, 2024, doi: 10.55123/insologi.v3i6.4627.
- [2] P. Global and D. D. A. N. Antisipasinya, "Oleh AGNES SRI MULYANI," 2021.
- [3] A. Frananda, A. Naibaho, and A. Zahra, "PREDIKSI KELULUSAN SISWA SEKOLAH MENENGAH PERTAMA MENGGUNAKAN MACHINE LEARNING," vol. 11, no. 3, 2023.
- [4] A. Khormarudin, "Data Mining Technique: K-Means Klustering Algorithm," *J. Ilmu Komput.*, pp. 1–12, 2016, [Online]. Available: <https://ilmukomputer.org/category/datamining/>
- [5] N. Krisdianto, R. Kyandra Atharizqi, W. T. Nurhasanaah, P. Informatika, U. Sultan, and A. Tirtayasa, "International Jurnal of Informatics Pengembangan Sistem Pengenalan Objek Multi-Kelas Berbasis Segmentasi Citra dengan YOLOv11 dan Streamlit," *IJInf Int. J. Informatics*, vol. 5, pp. 1–18, 2025.
- [6] P. N. Lhokseumawe, K. Pengantar, rahayu deny danar dan alvi furwanti Alwie, A. B. Prasetyo, and R. Andespa, *Implementasi Metode K-Means Klustering Untuk Pengelompokan Generasi Milenial Berdasarkan Perilaku Hemat Energi*, vol. 2, no. 1. 2020.
- [7] R. Pratama and L. Parinduri, "PENAGGULANGAN PEMANASAN GLOBAL Riza Pratama, Luthfi Parinduri," vol. 3814, pp. 91–95.
- [8] F. Febriansyah, "Penerapan Algoritma K-Means Klustering Data Gizi Balita Pada Uptd Puskesmas Bumi Agung," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4923.
- [9] G. A. B. Suryanegara, Adiwijaya, and M. D. Purbolaksono, "Improved Classification Results in the Random Forest Algorithm for Detection of Diabetes Patients Using the Normalization Method," *J. RESTI*, vol. 5, no. 1, pp. 114–122, 2021, doi: 10.29207/resti.v5i1.2880.
- [10] A. P. Margaretha, N. Ulinuha, and P. K. Intan, "Klustering Data Kecelakaan Lalu Lintas melalui Algoritma K-Means dengan Seleksi Fitur Chi-Square," pp. 215–224, 2024.
- [11] Y. D. Prasetyo, A. F. Hidayati, E. Maulida, and F. Y. Silalahi, "Implementasi Algoritma K-Means Klustering dalam Penentuan Gangguan Tidur Seseorang berdasarkan Gaya Hidup," *J. Inf. Eng. Educ. Technol.*, vol. 8, no. 1, pp. 54–61, 2024, doi: 10.26740/jieet.v8n1.p54-61.
- [12] A. W. Fuadah, F. N. Arifin, and O. Juwita, "Optimasi K-Klasterisasi Ketahanan Pangan Kabupaten Jember Menggunakan Metode Elbow," *INFORMAL Informatics J.*, vol. 6, no. 3, p. 136, 2021, doi: 10.19184/isj.v6i3.28363.
- [13] G. Sumantri, M. D. Novianto, and P. P. Prihastuti, "Implementasi Fuzzy C-Means dalam Pengelompokan Provinsi di Indonesia untuk Pemerataan Kualitas Pendidikan," *Pros. Semin. Pendidik. Mat. dan Mat.*, vol. 8, no. 2721, 2023, doi: 10.21831/pspmm.v8i2.310.
- [14] N. laras Agustina, *No Title*. 2019.
- [15] A. Salam, D. Adiatma, and J. Zeniarja, "Implementasi Algoritma K-Means Dalam



- Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS,” *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 62–68, 2020, doi: 10.33633/joins.v5i1.3350.
- [16] P. Bintoro, Ratnasari, E. Wihardjo, I. P. Putri, dan A. Asari, *Pengantar Machine Learning*. Solok: PT MAFY Media Literasi Indonesia, 2024. ISBN: 978-623-8758-78-4.