



Klasterisasi Kota di California Menggunakan *K-Means* untuk Analisis Geospasial Kebijakan Publik

Arfiannisa Kayla¹, Azay Agustian², Juan Montoya Darmawan³,
Revrigo Ramadhino⁴, Zurnan Alfian⁵

Program Studi Teknik Informatika Universitas Pamulang, Jl. Raya
Puspitek, Buaran, Kec. Pamulang, Kota Tangerang Selatan, Banten 15310

Abstrak

Received: 02 Januari 2026
Revised: 16 Januari 2026
Accepted: 28 Januari 2026

Penelitian ini bertujuan untuk mengelompokkan kota-kota di negara bagian California, Amerika Serikat, berdasarkan atribut geografis menggunakan algoritma *K-Means Clustering*. *Dataset* yang digunakan berisi 450 kota dengan fitur *latitude*, *longitude*, elevasi, dan jumlah populasi. Proses klasterisasi dilakukan setelah tahap *preprocessing* menggunakan *StandardScaler* untuk menyamakan skala data. Penentuan jumlah klaster optimal dilakukan dengan *metode Elbow* dan hasil evaluasi menggunakan *Silhouette Score* sebesar 0.5763 menunjukkan kualitas klaster yang cukup baik.

Selain *dataset* utama, penelitian ini juga membandingkan penerapan algoritma yang sama pada tiga *dataset* pendukung dari bidang sosial ekonomi, yaitu konsumsi pangan lokal di Indonesia, PDB per kapita negara ASEAN, dan distribusi kelas ekonomi masyarakat Indonesia. Hasil klasterisasi menunjukkan bahwa algoritma *K-Means* dapat digunakan secara fleksibel untuk berbagai jenis data dan dapat mendukung pengambilan kebijakan berbasis data, baik dalam perencanaan wilayah, ketahanan pangan, maupun pembangunan ekonomi regional.

Kata Kunci: *K-Means*, Klasterisasi, Data Geospasial, PDB ASEAN, Ketahanan Pangan,

(*) Corresponding Author: irzanrico@gmail.com, dosen02678@unpam.ac.id

How to Cite: Kayla, A., Agustian, A., Montoya Darmawan, J., Ramadhino, R., & Alfian, Z. (2026). Klasterisasi Kota di California Menggunakan *K-Means* untuk Analisis Geospasial Kebijakan Publik. *Jurnal Ilmiah Wahana Pendidikan*, 12(2.A), 113-127. Retrieved from <https://jurnal.peneliti.net/index.php/JIWP/article/view/12450>

PENDAHULUAN

Latar Belakang

Klasterisasi yaitu teknik analisis data yang tujuannya mengelompokkan objek-objek ke dalam kelompok berdasarkan kemiripan karakteristik tertentu. Dalam dunia nyata, teknik ini banyak digunakan dalam berbagai bidang, seperti pemasaran, pendidikan, *geospasial*, hingga perencanaan wilayah. Salah satu algoritma yang paling banyak digunakan dalam klasterisasi adalah *K-Means Clustering*, karena kesederhanaan dan efisiensi dalam menangani data berskala besar.

Dalam konteks perencanaan wilayah, pemetaan geografis yang efisien sangatlah penting, khususnya untuk wilayah dengan kepadatan kota tinggi seperti negara bagian California di Amerika Serikat. Klasterisasi berdasarkan lokasi kota (*latitude* dan *longitude*), *elevasi*, serta jumlah penduduk dapat membantu dalam merancang

strategi manajemen bencana, distribusi logistik, dan juga pembangunan infrastruktur yang lebih merata dan efisien.

Penelitian ini menggunakan *dataset* kota-kota di California yang bersumber dari *Python Data Science Handbook*, mencakup 450 kota dengan fitur numerik seperti *latitude*, *longitude*, *elevasi*, dan populasi total. Algoritma *K-Means* diterapkan untuk menemukan pola pengelompokan alami dalam data geografis tersebut. Hasil dari pengelompokan ini kemudian dianalisis untuk memahami bagaimana setiap *cluster* dapat berkontribusi dalam pengambilan keputusan berbasis wilayah.

Untuk memperkuat kajian dan memperluas konteks aplikatif dari metode klasterisasi ini, digunakan pula tiga *dataset* pendukung yang berasal dari domain sosial ekonomi, yaitu: konsumsi pangan karbohidrat lokal di Indonesia, data *PDB per kapita* negara ASEAN, dan distribusi kelas ekonomi di Indonesia. Ketiga *dataset* ini dianalisis menggunakan metode yang sama, dan hasilnya dibandingkan dengan *dataset* utama untuk menunjukkan fleksibilitas dan kekuatan analisis *clustering* dalam berbagai bidang.

TINJAUAN PUSTAKA

Data mining merupakan proses untuk mengekstraksi pola atau informasi yang bermakna dari kumpulan data yang besar dan kompleks. Salah satu teknik penting dalam *data mining* adalah *clustering*, yaitu metode pembelajaran tanpa pengawasan (*unsupervised learning*) yang bertujuan mengelompokkan data tanpa label berdasarkan kemiripan karakteristik atau fitur. *Clustering* memudahkan analisis data dengan mengelompokkan objek-objek yang memiliki kesamaan ke dalam satu kelompok, sehingga pola tersembunyi dapat diidentifikasi secara lebih efisien.

Salah satu algoritma *clustering* yang paling populer dan banyak digunakan adalah *K-Means*. *K-Means* merupakan algoritma berbasis *centroid* yang bekerja dengan membagi data ke dalam sejumlah klaster yang telah ditentukan sebelumnya (K), kemudian menghitung jarak antara setiap data dengan *centroid cluster* untuk menentukan keanggotaan klaster. Proses ini dilakukan secara iteratif dengan memperbarui posisi *centroid* hingga tidak terjadi perubahan signifikan pada keanggotaan klaster. Kelebihan utama *K-Means* adalah kesederhanaan dan efisiensinya dalam menangani data numerik berskala besar.

Penentuan jumlah *cluster* yang optimal sangat penting dalam proses *clustering*. Salah satu metode yang sering digunakan adalah *Elbow Method*, yaitu dengan memplot variasi yang dijelaskan sebagai fungsi dari jumlah klaster dan memilih titik “*elbow*” pada grafik sebagai jumlah *cluster* yang optimal. Titik *elbow* menandai perubahan signifikan pada penurunan nilai *inertia*, sehingga menambah klaster setelah titik tersebut tidak memberikan peningkatan yang berarti dalam pemodelan data.

Selain itu, *Silhouette Score* juga digunakan untuk mengevaluasi kualitas hasil *clusterisasi*. *Silhouette Score* mengukur seberapa baik setiap data cocok dengan klaster-nya sendiri dibandingkan dengan klaster lain, dengan nilai berkisar antara -1 hingga 1. Nilai mendekati 1 menunjukkan klaster yang baik dan terpisah jelas, sedangkan nilai mendekati 0 menunjukkan klaster yang tumpang tindih. Penggunaan *Silhouette Score* dapat membantu dalam membandingkan performa beberapa algoritma *clustering* atau menentukan jumlah *cluster* terbaik untuk *dataset* tertentu.

Dalam berbagai bidang, *K-Means* telah diaplikasikan secara luas. Pada studi *geospasial*, algoritma ini digunakan untuk segmentasi wilayah geografis berdasarkan data spasial dan atribut lainnya, seperti yang dilakukan pada pemodelan lingkungan akustik luar ruang.

Dalam penelitian ini, *K-Means* diterapkan tidak hanya pada data geografis, tetapi juga pada data sosial ekonomi seperti konsumsi pangan, pertumbuhan ekonomi regional, dan distribusi kelas ekonomi. Hal ini bertujuan untuk menunjukkan generalisasi metode *K-Means* dalam mendukung analisis dan perencanaan kebijakan publik di berbagai sektor.

RUMUSAN MASALAH

Berdasarkan latar belakang dan tinjauan pustaka yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana algoritma *K-Means* digunakan untuk mengelompokkan kota-kota di California berdasarkan karakteristik geografis seperti *latitude*, *longitude*, *elevasi*, dan populasi?
2. Berapa jumlah *cluster* optimal yang dapat dibentuk berdasarkan hasil evaluasi menggunakan metode *Elbow* dan *Silhouette Score*?
3. Bagaimana perbandingan hasil *clustering* pada *dataset* geografis dengan hasil klusterisasi dari *dataset* sosial ekonomi (konsumsi pangan lokal, PDB ASEAN, dan distribusi kelas ekonomi Indonesia)?
4. Apa saja potensi pemanfaatan hasil *clustering* dalam mendukung perencanaan kebijakan publik yang efektif dan berbasis data?

METODE PENELITIAN

1. Tools dan Teknologi

Pengolahan data dilakukan menggunakan bahasa pemrograman *Python* dengan dukungan berbagai pustaka (*library*) populer yang umum digunakan dalam analisis data dan pembelajaran mesin. Proses pengembangan dilakukan pada aplikasi editor *Visual Studio Code (VSCode)*. Berikut adalah pustaka dan teknologi yang digunakan:

- *Pandas*
Merupakan pustaka *Python* yang digunakan untuk manipulasi dan analisis data berbasis tabel, memungkinkan pengolahan data secara efisien dalam format *DataFrame* (McKinney, 2018).
- *NumPy*
Digunakan untuk komputasi numerik dengan *array* multidimensi serta menyediakan fungsi matematis tingkat lanjut yang sangat berguna dalam pengolahan data numerik (Oliphant, 2015).
- *Matplotlib*
Merupakan pustaka visualisasi data yang memungkinkan pembuatan berbagai jenis grafik seperti garis, batang, dan *scatter plot* untuk eksplorasi data secara visual (Hunter, 2007).
- *Scikit-Learn (sklearn)*
Pustaka utama untuk pembelajaran mesin di *Python*, digunakan untuk *preprocessing data*, implementasi algoritma *K-Means*, serta evaluasi model *clustering* secara efektif (Pedregosa et al., 2011).

- *Visual Studio Code (VSCode)*

Digunakan sebagai lingkungan pengembangan terintegrasi (IDE) yang mendukung penulisan, eksekusi, dan debugging kode *Python* secara efisien, mempermudah proses pengembangan aplikasi *data mining* (Ningsih et al., 2022).

Penggunaan *Python* dan pustaka-pustaka tersebut memungkinkan analisis data dan penerapan algoritma machine learning berjalan secara efektif dan efisien, sesuai dengan perkembangan teknologi *data mining* terkini (Research Consultant, 2023).

2. Sumber Data

Data yang digunakan dalam penelitian ini terdiri dari satu *dataset* utama dan tiga *dataset* pendukung. *Dataset* utama diperoleh dari repositori terbuka yang digunakan dalam *Python Data Science Handbook*, sedangkan *dataset* pendukung berasal dari hasil penyusunan mahasiswa berdasarkan data resmi.

Dataset utama berisi informasi 450 kota di negara bagian California, Amerika Serikat, dengan atribut berupa koordinat geografis (*latitude* dan *longitude*), ketinggian wilayah (*elevation*), jumlah penduduk, dan nama kota. Data ini diambil dari file *california_cities.csv* yang dapat diakses secara terbuka melalui *GitHub*.

Dataset pendukung pertama berisi data konsumsi pangan karbohidrat lokal di Indonesia berdasarkan kontribusi energi, berat setara beras, dan kenaikan konsumsi per kapita. *Dataset* kedua memuat nilai Produk Domestik Bruto (PDB) per kapita dari negara-negara ASEAN pada tahun 2006 hingga 2008. Sementara *dataset* ketiga menampilkan distribusi penduduk berdasarkan kelas ekonomi di Indonesia pada tahun 1999 dan 2009.

Semua data digunakan sebagai dasar untuk proses analisis kluster menggunakan algoritma *K-Means* guna mencari pola pengelompokan berdasarkan kemiripan fitur numerik.

3. Preprocessing Data

Tahapan *preprocessing* data dilakukan sebelum proses pengelompokan (*clustering*) agar data yang digunakan berada dalam skala yang seragam dan sesuai dengan kebutuhan algoritma. *Dataset* utama yang digunakan adalah *california_cities.csv*, yang memuat informasi mengenai 450 kota di negara bagian California, Amerika Serikat.

Langkah pertama adalah memilih fitur yang relevan untuk analisis kluster, yaitu *latitude* (*latd*) dan *longitude* (*longd*), karena penelitian ini berfokus pada pengelompokan geografis. Selanjutnya, dilakukan proses standarisasi menggunakan metode *StandardScaler* dari pustaka *sklearn.preprocessing*. Proses ini mengubah skala data menjadi distribusi standar dengan rata-rata 0 dan standar deviasi 1, untuk memastikan bahwa tidak ada fitur yang mendominasi proses perhitungan jarak antar titik.

Tahapan *preprocessing* ini penting karena algoritma *K-Means* sangat sensitif terhadap perbedaan skala antar fitur. Setelah data distandarisasi, proses klusterisasi dapat dilakukan secara optimal.

4. Algoritma

Algoritma yang digunakan dalam penelitian ini adalah *K-Means Clustering*, yaitu metode *unsupervised learning* yang bertujuan mengelompokkan data ke dalam sejumlah kluster berdasarkan kemiripan fitur numerik. *K-Means* bekerja

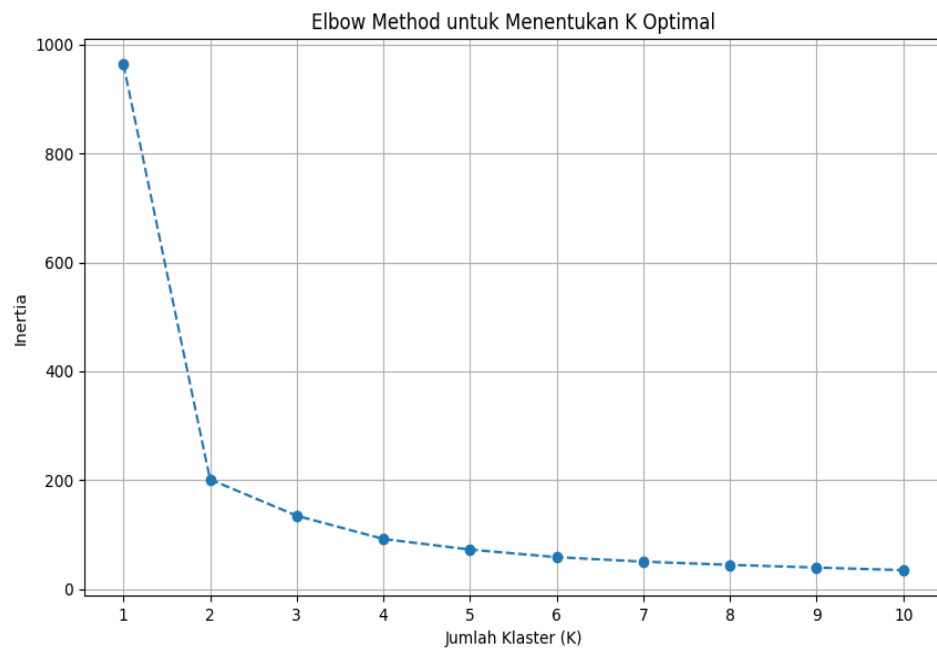
dengan menentukan sejumlah pusat kluster (*centroid*), lalu mengelompokkan data berdasarkan jarak terdekat ke *centroid* tersebut, menggunakan perhitungan *jarak Euclidean*.

Langkah-langkah algoritma *K-Means*:

1. Menentukan jumlah kluster *K*.
2. Menginisialisasi pusat kluster (*centroid*) secara acak.
3. Menghitung jarak setiap data ke seluruh *centroid*.
4. Menempatkan data ke kluster terdekat.
5. Menghitung ulang posisi *centroid* berdasarkan rata-rata anggota kluster.
6. Mengulangi proses hingga hasil *konvergen* (tidak ada perubahan kluster).

Implementasi Kode *Python*:

```
1 # --- MENENTUKAN JUMLAH KLASTER (K) OPTIMAL DENGAN ELBOW METHOD ---
2
3 # 5. Hitung "Within-Cluster Sum of Squares" (Inertia) untuk berbagai nilai K
4 inertia = []
5 K_range = range(1, 11)
6 for k in K_range:
7     kmeans_temp = KMeans(n_clusters=k, random_state=42, n_init=10)
8     kmeans_temp.fit(features_scaled)
9     inertia.append(kmeans_temp.inertia_)
10
11 # 6. Visualisasi Elbow Method
12 plt.figure(figsize=(10, 6))
13 plt.plot(K_range, inertia, marker='o', linestyle='--')
14 plt.title('Elbow Method untuk Menentukan K Optimal')
15 plt.xlabel('Jumlah Klaster (K)')
16 plt.ylabel('Inertia')
17 plt.xticks(K_range)
18 plt.grid(True)
19 print("--- Menampilkan Plot Elbow Method (Tutup untuk melanjutkan) ---")
20 plt.show()
```



Penentuan Jumlah Kluster (K):

Untuk menentukan jumlah kluster yang optimal, digunakan *Elbow Method*, yaitu teknik visualisasi nilai *inertia* untuk berbagai nilai K . Titik di mana grafik mulai melandai dianggap sebagai jumlah *cluster* terbaik. Berdasarkan hasil visualisasi, dipilih nilai $K = 4$, yang dianggap memberikan keseimbangan terbaik antara kompleksitas model dan hasil segmentasi.

5. Evaluasi

Setelah proses klusterisasi dilakukan menggunakan algoritma *K-Means* dengan $K = 4$, langkah selanjutnya adalah melakukan evaluasi terhadap kualitas kluster yang terbentuk. Evaluasi dilakukan menggunakan *Silhouette Score*, yaitu metrik yang mengukur seberapa baik data cocok berada dalam kluster tertentu dan seberapa jauh data tersebut dari kluster lainnya.

Hasil Evaluasi Model:

```
1 # 8. Evaluasi Model dengan Silhouette Score
2 # Skor Silhouette mengukur seberapa mirip sebuah objek dengan klusternya sendiri dibandingkan dengan kluster lain.
3 # Rentang skor: -1 (salah kluster) hingga +1 (kluster sangat padat). 0 menunjukkan kluster tumpang tindih.
4 score = silhouette_score(features_scaled, df['cluster'])
5 print(f"\n--- Evaluasi Model ---")
6 print(f"Jumlah Kluster (K) yang Digunakan: {optimal_k}")
7 print(f"Silhouette Score: {score:.4f}")
```

```
--- Evaluasi Model ---
Jumlah Kluster (K) yang Digunakan: 4
Silhouette Score: 0.5763
(Skor mendekati +1 menunjukkan kluster yang baik dan terpisah)
```

Berdasarkan hasil perhitungan, diperoleh nilai *Silhouette Score* sebesar 0.5763, yang menunjukkan bahwa hasil kluster yang terbentuk berada pada kategori cukup baik. Kluster-kluster yang dihasilkan cukup terpisah satu sama lain dan menunjukkan struktur yang dapat diinterpretasikan secara geografis. Evaluasi ini memperkuat pemilihan nilai $K = 4$ sebagai jumlah kluster optimal berdasarkan *Elbow Method* yang sebelumnya dilakukan.

6. Dataset

Dataset yang digunakan dalam penelitian ini adalah data kota-kota di negara bagian California, Amerika Serikat, yang diperoleh dari buku *Python Data Science Handbook* karya Jake VanderPlas. *Dataset* tersebut tersedia dalam format .csv dengan nama *california_cities.csv*, dan berisi informasi geografis dan demografis dari 450 kota.

Struktur *Dataset*:

Nama Kolom	Tipe Data	Deskripsi
<i>city</i>	<i>String</i>	Nama kota
<i>latd</i>	<i>Float</i>	Koordinat lintang (<i>latitude</i>) kota
<i>longd</i>	<i>Float</i>	Koordinat bujur (<i>longitude</i>) kota
<i>elevation_m</i>	<i>Float</i>	Ketinggian kota dari permukaan laut dalam meter
<i>population_total</i>	<i>Integer</i>	Jumlah penduduk kota

Fitur yang Digunakan:

Dalam penelitian ini, fitur yang digunakan untuk proses klusterisasi hanya mencakup:

- *latd (latitude)*
- *longd (longitude)*

Fitur tersebut dipilih karena tujuan utama dari penelitian ini adalah melakukan klusterisasi spasial (geografis), untuk mengelompokkan kota-kota berdasarkan lokasi koordinatnya.

7. Pengolahan Data

Proses pengolahan data dalam penelitian ini dilakukan melalui beberapa tahapan utama, mulai dari pemilihan fitur hingga visualisasi hasil kluster. Semua proses dilakukan menggunakan bahasa pemrograman *Python* dengan pustaka pendukung seperti *Pandas*, *Scikit-Learn*, dan *Matplotlib*. Berikut tahapan pengolahan data yang dilakukan:

1. Pemilihan Fitur (*Data Selection*):

Dataset california_cities.csv memiliki beberapa kolom, namun hanya dua fitur numerik yang digunakan untuk proses klusterisasi, yaitu *latd (latitude)* dan *longd*

(*longitude*), karena fokus penelitian adalah klasterisasi spasial berdasarkan lokasi geografis.

2. Standarisasi (*Data Preprocessing*):

Data yang dipilih distandarisasi menggunakan *StandardScaler* agar memiliki rata-rata 0 dan standar deviasi 1. Hal ini dilakukan untuk memastikan bahwa skala antar fitur tidak mempengaruhi proses perhitungan jarak dalam algoritma *K-Means*.

3. Penentuan Jumlah Klaster (*Elbow Method*):

Untuk menentukan jumlah klaster optimal, digunakan metode *Elbow*, yaitu dengan menghitung nilai *inertia* (jumlah kuadrat jarak titik ke *centroid*) untuk $K = 1$ hingga 10. Hasil visualisasi menunjukkan bahwa nilai K optimal berada pada $K = 4$, karena terjadi “tekukan” (*elbow*) yang signifikan.

4. Pelatihan Model *K-Means*:

Model *K-Means* diterapkan pada data yang telah distandarisasi dengan parameter $n_clusters=4$. Proses ini menghasilkan klaster dan memberikan label klaster untuk setiap kota.

5. Evaluasi Model:

Evaluasi dilakukan menggunakan *Silhouette Score*, dan diperoleh skor sebesar 0.5763, yang menunjukkan bahwa hasil klaster cukup baik dan klaster-klaster terbentuk dengan pemisahan yang jelas.

6. Visualisasi Hasil Klastering:

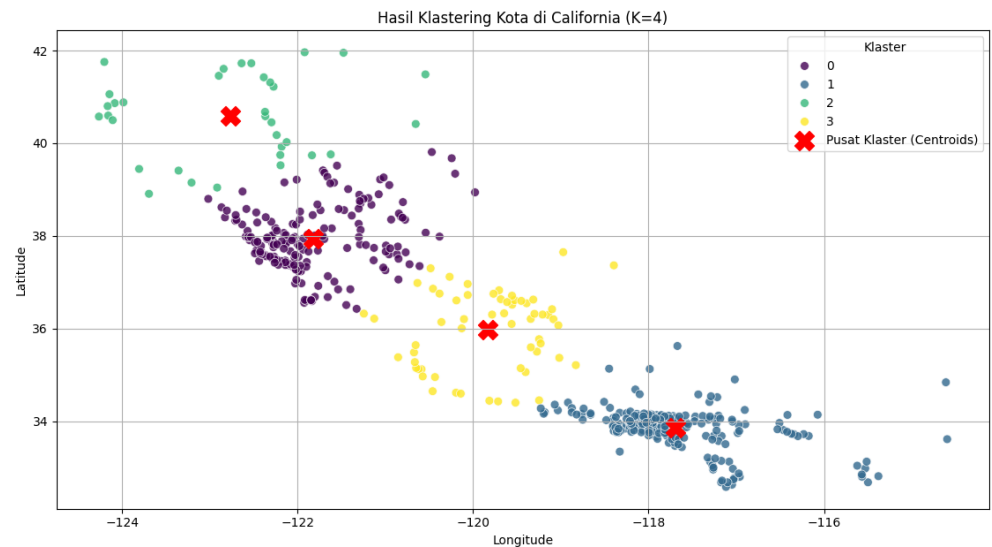
Hasil *clustering* divisualisasikan menggunakan *scatter plot* dengan koordinat *longitude* dan *latitude*, dan dibedakan berdasarkan warna klaster. Pusat klaster (*centroid*) juga ditampilkan untuk menunjukkan titik pusat dari masing-masing klaster.

HASIL DAN PEMBAHASAN

Hasil Klasterisasi *Dataset* California

Proses klasterisasi menggunakan algoritma *K-Means* dengan jumlah klaster $K = 4$ menghasilkan pengelompokan geografis kota-kota di California yang cukup jelas. Setiap kota diberi label klaster 0, 1, 2, atau 3 berdasarkan kedekatan koordinat lokasinya (*latitude* dan *longitude*).

Visualisasi hasil *clustering* ditampilkan pada *scatter plot* dua dimensi yang memetakan posisi kota berdasarkan bujur (*longitude*) dan lintang (*latitude*). Warna pada grafik merepresentasikan klaster yang berbeda, sementara titik berwarna merah berbentuk silang (X) menandai posisi pusat klaster (*centroid*).



Dari hasil visualisasi, dapat dilihat bahwa:

- Klaster 0 mencakup wilayah pesisir barat bagian utara.
- Klaster 1 mencakup wilayah tengah dan bagian selatan.
- Klaster 2 tersebar di area timur laut dan dataran tinggi.
- Klaster 3 terletak di daerah yang lebih terpencil atau padat tertentu.

Hal ini menunjukkan bahwa algoritma *K-Means* mampu mengelompokkan kota-kota berdasarkan kedekatan spasial, yang sangat berguna dalam konteks pemetaan infrastruktur, manajemen risiko bencana, serta perencanaan logistik wilayah.

Silhouette Score sebesar 0.5763 mengindikasikan bahwa hasil klaster cukup baik, di mana jarak antar klaster cukup jauh dan data dalam klaster memiliki kesamaan tinggi.

Komparasi dengan *Dataset* Pendukung

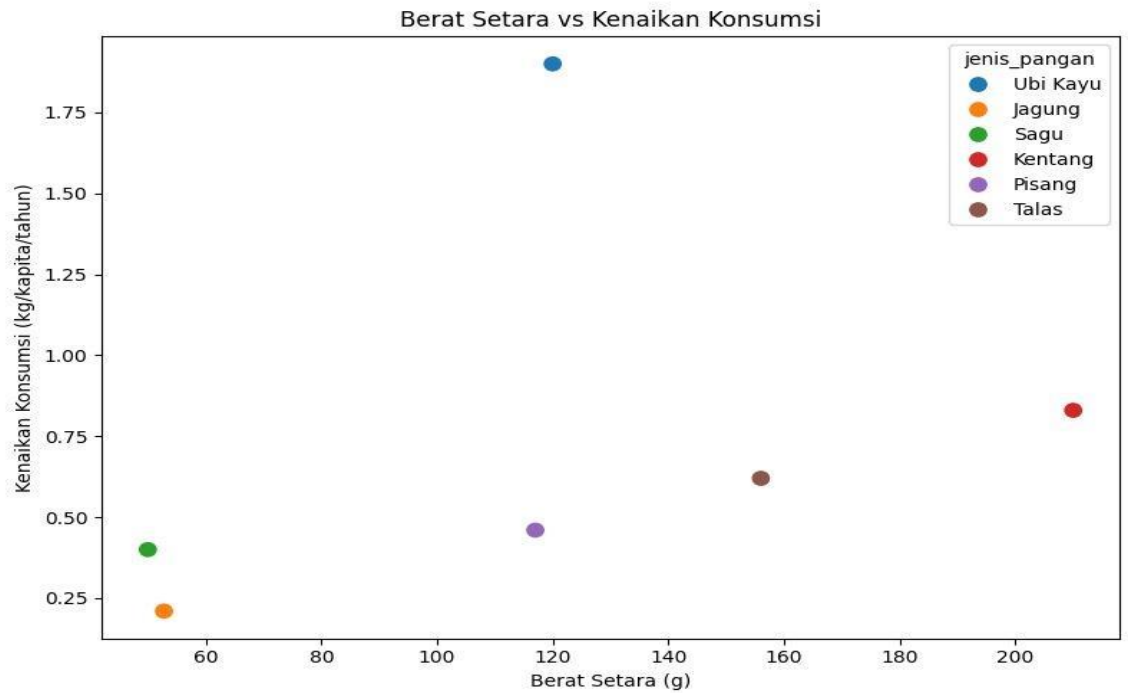
Untuk memperluas konteks aplikasi algoritma *K-Means*, digunakan tiga *dataset* tambahan yang berasal dari bidang sosial ekonomi. Berikut ringkasan temuan dan perbandingannya:

1. *Dataset* Konsumsi Pangan Lokal

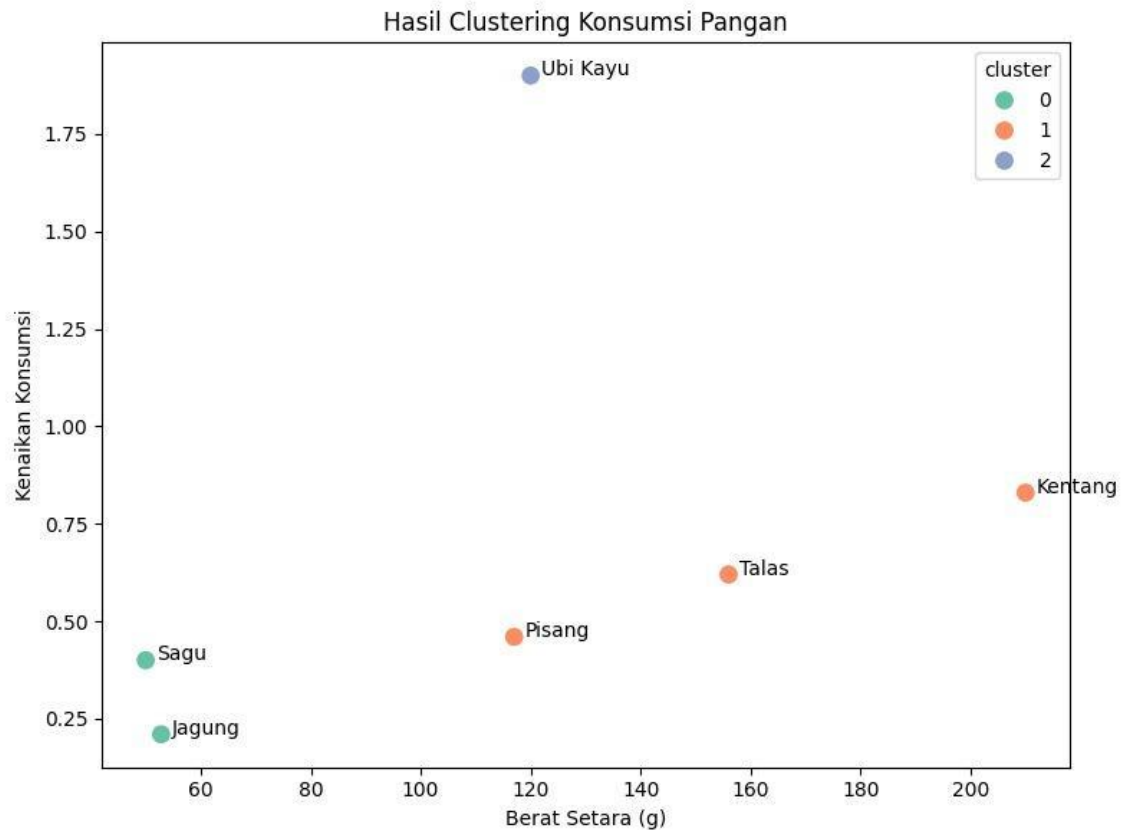
Dataset ini memuat enam jenis pangan lokal Indonesia, yaitu: ubi kayu, jagung, sagu, kentang, pisang, dan talas. Setiap jenis pangan memiliki dua variabel utama:

- Berat setara beras (gram)
- Kenaikan konsumsi (kg/kapita/tahun)

Tujuan klasterisasi pada *dataset* ini adalah untuk menemukan pola dalam preferensi dan potensi pangan lokal sebagai sumber energi alternatif selain beras. Dengan menggunakan algoritma *K-Means*, data dikelompokkan berdasarkan kesamaan nilai pada dua variabel tersebut.



Visualisasi ini menunjukkan distribusi awal dari masing-masing jenis pangan berdasarkan dua variabel. Ubi kayu tampak memiliki kontribusi konsumsi tertinggi, sedangkan jagung dan sagu berada di area bawah grafik dengan kontribusi yang lebih rendah.



Setelah dilakukan klasterisasi, terbentuk tiga klaster utama:

- Klaster 0 (hijau): Pangan dengan kontribusi rendah (sagu dan jagung)
- Klaster 1 (oranye): Pangan sedang (pisang dan talas)
- Klaster 2 (ungu/biru): Pangan dengan kontribusi sangat tinggi (ubi kayu)

Interpretasi:

Klasterisasi ini memperlihatkan potensi **ubi kayu** sebagai kandidat utama dalam program diversifikasi pangan nasional, karena menempati klaster tersendiri dengan tingkat kenaikan konsumsi dan berat setara beras yang paling tinggi. Informasi ini dapat digunakan oleh pemerintah dalam menetapkan prioritas dukungan terhadap produksi dan distribusi pangan lokal.

Relevansi terhadap Kebijakan Publik:

Hasil klasterisasi ini dapat digunakan sebagai dasar pertimbangan dalam pengambilan kebijakan ketahanan pangan nasional, khususnya terkait program diversifikasi pangan yang bertujuan mengurangi ketergantungan masyarakat terhadap beras sebagai sumber utama karbohidrat.

Dengan diketahuinya klaster pangan yang memiliki kontribusi tinggi terhadap konsumsi (seperti ubi kayu), pemerintah daerah atau pusat dapat:

- Mengarahkan subsidi dan pelatihan pertanian ke jenis pangan unggulan di tiap klaster.
- Menyesuaikan distribusi logistik dan rantai pasok pangan secara lebih efisien berdasarkan karakteristik konsumsinya.
- Merancang kampanye promosi konsumsi pangan lokal berdasarkan potensi masing-masing klaster.

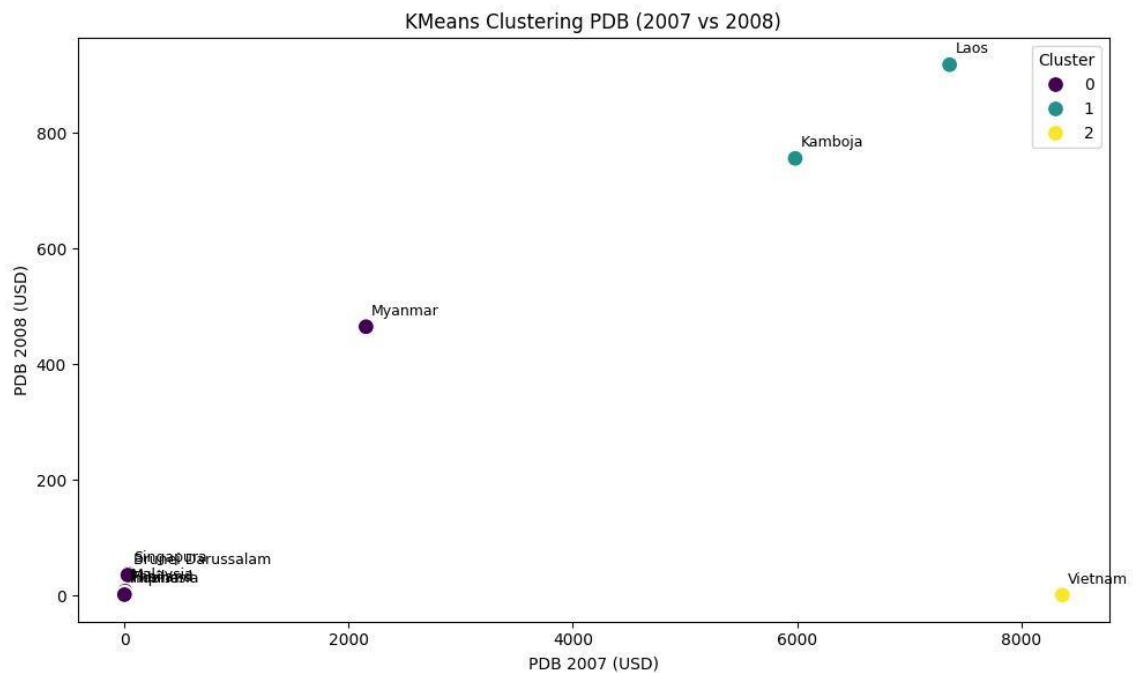
Dengan demikian, pendekatan klasterisasi ini tidak hanya membantu dalam analisis data, tetapi juga memiliki dampak nyata dalam menyusun strategi pangan yang berbasis data dan kontekstual.

2. *Dataset* PDB ASEAN

Dataset ini memuat data Produk Domestik Bruto (PDB) per kapita dalam USD dari negara-negara anggota ASEAN pada tahun 2007 dan 2008. Tujuan klasterisasi adalah untuk melihat pengelompokan negara berdasarkan kemiripan tingkat pertumbuhan ekonomi, yang berguna dalam mengidentifikasi pola perkembangan ekonomi kawasan.

Fitur yang Digunakan:

- PDB per kapita tahun 2007
- PDB per kapita tahun 2008



Interpretasi:

Dari hasil klasterisasi, terbentuk **tiga klaster** yang menggambarkan kondisi ekonomi negara-negara ASEAN secara umum:

- Klaster 0 (Ungu)

Negara-negara yang termasuk dalam klaster ini (Indonesia, Filipina, Malaysia, Brunei, Singapura) memiliki tingkat PDB per kapita relatif rendah hingga menengah, namun cenderung stabil dan saling berdekatan secara ekonomis. Beberapa negara di sini memiliki PDB tinggi namun tidak mengalami lonjakan signifikan.

- Klaster 1 (Biru Kehijauan)

Terdiri dari Laos dan Kamboja, yang memiliki PDB menengah dan menunjukkan pertumbuhan yang cukup pesat dalam dua tahun tersebut. Klaster ini merepresentasikan negara-negara dengan potensi ekonomi berkembang namun belum setara dengan negara-negara di klaster 0.

- Klaster 2 (Kuning)

Dihuni oleh Vietnam, yang berada di posisi PDB tinggi pada 2007 tetapi tidak menunjukkan pertumbuhan besar di 2008. Negara ini membentuk klaster tersendiri, menandakan pola pertumbuhan yang berbeda dan menonjol dari negara lain.

Relevansi terhadap Kebijakan Ekonomi Regional:

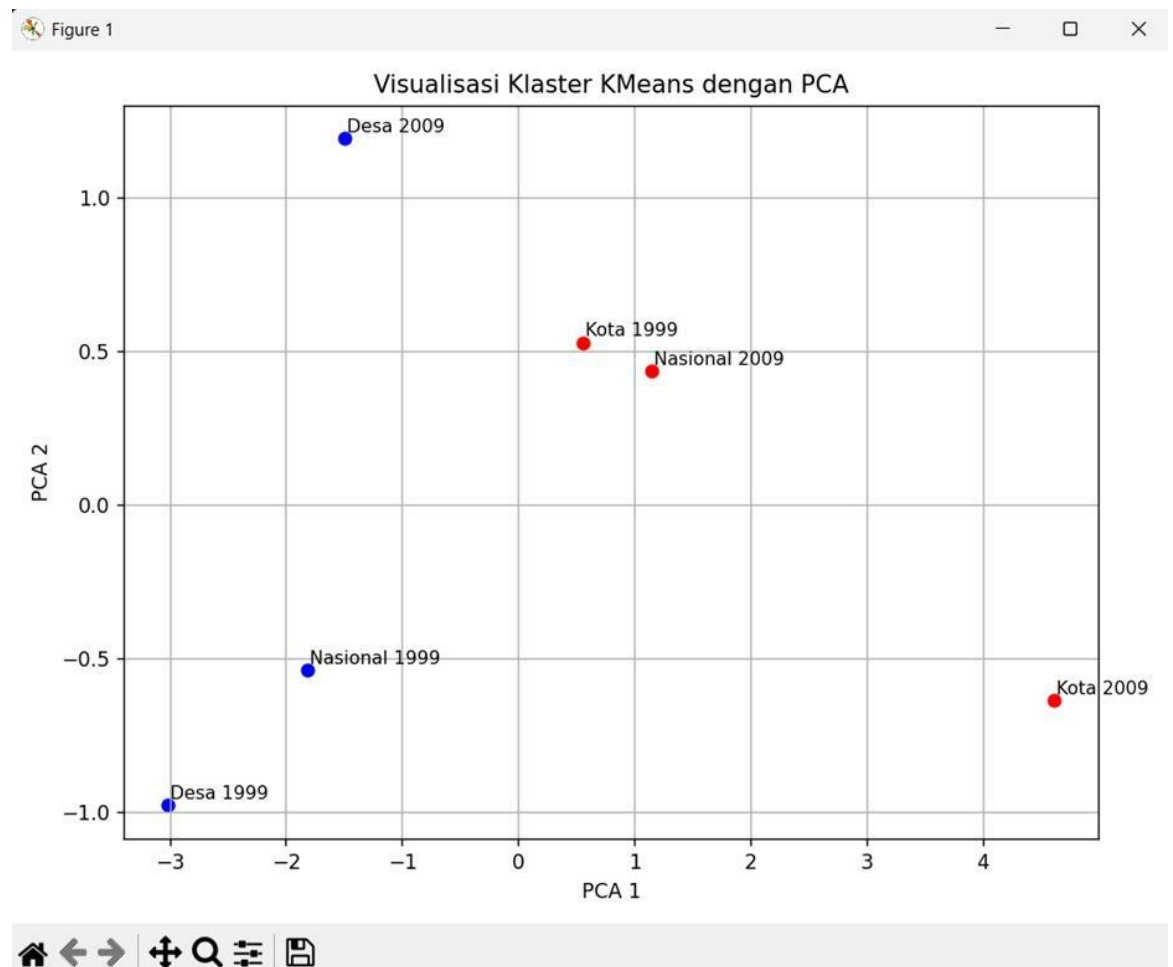
- Pemerintah dan lembaga ASEAN dapat menggunakan hasil klaster ini untuk menyusun program regional yang lebih tersegmentasi, seperti pembiayaan infrastruktur, bantuan teknis, atau pengembangan teknologi.
- Negara-negara dalam satu klaster dapat dijadikan mitra strategis dalam pengembangan bersama, karena kesamaan tingkat ekonomi.
- Analisis ini juga menunjukkan pentingnya penggunaan data kuantitatif untuk memahami struktur ekonomi kawasan yang tidak selalu tercermin dalam keanggotaan organisasi semata.

3. *Dataset* Distribusi Kelas Ekonomi

Dataset ini menyajikan distribusi kelas ekonomi masyarakat Indonesia pada dua periode, yaitu tahun 1999 dan 2009, berdasarkan data pengeluaran per kapita. Data diklasifikasikan menjadi tiga wilayah: nasional, kota, dan desa. Tujuan dari klusterisasi adalah untuk melihat perubahan pola distribusi ekonomi selama satu dekade, serta mengelompokkan data berdasarkan kemiripan kondisi ekonomi antarkelompok wilayah dan waktu.

Fitur yang Digunakan:

- Jenis wilayah: nasional, kota, desa
- Tahun: 1999 dan 2009
- Persentase pengeluaran > USD 2 per kapita per hari



Interpretasi Kluster:

Berdasarkan hasil klusterisasi yang divisualisasikan menggunakan PCA *projection (Principal Component Analysis)*, diperoleh dua kluster utama yang membagi data berdasarkan tahun dan kategori wilayah:

- Kluster Merah (Tahun 1999):

Berisi data dari tahun 1999 yang menggambarkan kondisi awal ekonomi masyarakat. Kota dan nasional masih berada dalam satu kluster yang relatif homogen.

- **Klaster Biru (Tahun 2009):**

Berisi data dari tahun 2009 yang menunjukkan adanya pergeseran signifikan, terutama pada kategori kota, yang berpindah klaster dari posisi tahun 1999. Hal ini menunjukkan peningkatan jumlah masyarakat kelas menengah ke atas di wilayah perkotaan.

Klaster untuk desa tidak berubah secara signifikan, menunjukkan bahwa perkembangan ekonomi di wilayah desa cenderung stagnan atau lebih lambat dibandingkan kota.

Relevansi terhadap Kebijakan Sosial dan Pemerataan Ekonomi:

- Pemerintah dapat menggunakan pola klaster ini untuk menyusun program pengentasan kemiskinan yang lebih terarah berdasarkan wilayah dan tahun.
- Perubahan signifikan di wilayah kota menunjukkan perlunya perhatian pada kesenjangan desa-kota yang semakin melebar.
- Klasterisasi ini dapat menjadi bahan evaluasi kebijakan pembangunan ekonomi selama satu dekade, serta memperkuat argumentasi pentingnya perencanaan berbasis data dalam pembangunan sosial.

KESIMPULAN

Penelitian ini menerapkan algoritma *K-Means Clustering* pada *dataset* geografis kota-kota di California, Amerika Serikat, untuk membentuk klaster wilayah berbasis koordinat lintang dan bujur. Berdasarkan hasil evaluasi menggunakan *Elbow Method* dan *Silhouette Score* (0.5763), diperoleh empat klaster spasial yang merepresentasikan segmentasi wilayah yang jelas secara geografis. Hasil ini memiliki potensi untuk mendukung kebijakan publik seperti penempatan fasilitas, manajemen bencana, dan distribusi logistik yang lebih efektif.

Sebagai pembandingan, tiga *dataset* pendukung dianalisis dengan algoritma yang sama:

- *Dataset* konsumsi pangan lokal Indonesia menunjukkan potensi diversifikasi sumber karbohidrat alternatif.
- *Dataset* PDB ASEAN menunjukkan segmentasi pertumbuhan ekonomi yang relevan untuk kebijakan regional.
- *Dataset* distribusi kelas ekonomi Indonesia menampilkan pola pergeseran kesejahteraan masyarakat kota dan desa selama satu dekade.

Hasil ini membuktikan bahwa *K-Means* dapat digunakan secara fleksibel di berbagai domain, baik spasial maupun sosial-ekonomi. Ke depan, penelitian ini dapat dikembangkan dengan:

- Menambahkan variabel lain seperti indeks sosial, pendidikan, atau infrastruktur.
- Menggunakan algoritma klasterisasi alternatif seperti DBSCAN atau *Hierarchical Clustering* untuk hasil yang lebih adaptif.

DAFTAR PUSTAKA

Sharma, N. (2025). *K-Means Clustering Explained*. Diambil dari <https://neptune.ai/blog/k-means-clustering>

Elbow method (clustering). (n.d.). Diambil dari https://handwiki.org/wiki/Elbow_meth

Lee, S. (n.d.). *Mastering Silhouette Score for Effective Clustering*. Diambil dari <https://www.numberanalytics.com/blog/mastering-silhouette-score-clustering>

Pedersen, K., Jensen, R. R., Hall, L. K., Cutler, M. C., Transtrum, M. K., Gee, K. L., & Lympany, S. V. (2023). *K-Means Clustering of 51 Geospatial Layers Identified for Use in Continental-Scale Modeling of Outdoor Acoustic Environments*. Diambil dari <https://www.mdpi.com/2076-3417/13/14/8123>

Fa'rifah, R. Y., & Pramesti, D. (n.d.). *Cluster Analysis of Inclusive Economic Development Using K-Means Algorithm*. Diambil dari <https://journal.universitasbumigora.ac.id/in>

Miranti. (2023). *Penggunaan Python untuk Data Mining*. Diambil dari <https://researchconsultant.id/artikel/penggunaan-python-untuk-data-mining/>

Penggunaan Python untuk Data Mining. (n.d.). Diambil dari <https://socs.binus.ac.id/2018/11/16/penggunaan-python-untuk-data-mining/>

Ramadhino, R. (2025). Source code dan tangkapan layar. Diambil dari <https://drive.google.com/drive/folders/1gS8d-6BtYYdosbrb4IF80FDmYyR1naI3?usp=sharing>

Kayla, A. (2025). Source code dan tangkapan layar. Diambil dari <https://drive.google.com/drive/folders/1y5ACVx9oIUQi7z9uTkxC9PfAyTHCXsFg?usp=sharing>

Agustian, A. (2025). Source code dan tangkapan layar. Diambil dari https://drive.google.com/drive/folders/1H0EHS2RXBIWRvtkhIFKM_WUSsCIg-kJI?usp=sharing

Montoya, J. D. (2025). Source code dan tangkapan layar. Diambil dari https://drive.google.com/drive/folders/1JPtJ12ScnZAVuwSx0f_LCp8JoidyRqP?usp=sharing