

Perbandingan Hasil Klasterisasi K-Means Berdasarkan Silhouette Score dan Inertia Pada 5 Dataset Berbeda

Septiantoro Sogema Fourteen¹, Al Hakim Putra Bintang Anugrah², Muhamad Niko Aninda Putra³, Jihan Aulia Shabrina⁴, Mulyo Adji Winoto⁵, Zurnan Alfian⁶

^{1,2,3,4,5,6} Teknik Informatika, Universitas Pamulang

e-mail: toroseptiantoro@gmail.com

Abstrak

Clustering merupakan metode penting dalam data mining untuk mengelompokkan data berdasarkan kemiripan. K-Means adalah salah satu algoritma yang banyak digunakan karena kesederhanaannya dalam membagi data ke dalam sejumlah kluster. Penelitian ini membandingkan hasil klasterisasi algoritma K-Means yang dilakukan oleh lima mahasiswa menggunakan lima dataset yang berbeda, dengan variasi jumlah kluster dan tahapan preprocessing. Evaluasi dilakukan menggunakan metrik internal, yaitu Silhouette Score dan Inertia. Hasil menunjukkan adanya variasi signifikan dalam kualitas kluster yang dihasilkan, bergantung pada karakteristik data dan parameter yang digunakan. Nilai Silhouette Score tertinggi sebesar 0,6946 dan Inertia terendah sebesar 1,5 ditemukan pada dataset yang berbeda. Penelitian ini memberikan gambaran praktis mengenai pentingnya pemilihan parameter dan tahapan preprocessing dalam meningkatkan kualitas hasil clustering dan juga memberikan wawasan bagi pemula dalam memahami evaluasi hasil klasterisasi secara praktis.

Kata Kunci: *K-Means, Clustering, Silhouette Score, Inertia, Klasterisasi*

Abstract

Clustering is an important method in data mining to group data based on similarity. K-Means is one of the widely used algorithms because of its simplicity in dividing data into a number of clusters. This study compares the results of the K-Means algorithm clustering carried out by five students using five different datasets, with variations in the number of clusters and preprocessing stages. The evaluation was carried out using internal metrics, namely Silhouette Score and Inertia. The results show significant variations in the quality of the resulting clusters, depending on the characteristics of the data and the parameters used. The highest Silhouette Score value of 0.6946 and the lowest Inertia of 1.5 were found in different datasets. This study provides a practical overview of the importance of parameter selection and preprocessing stages in improving the quality of clustering results and also provides insight for beginners in understanding the evaluation of clustering results practically.

Keywords: *K-Means, Clustering, Silhouette Score, Inertia, clustering.*

PENDAHULUAN

Dalam Era Digital ini, kemampuan dalam menganalisis dan menggali informasi menjadi keterampilan paling krusial di berbagai bidang. Salah satu pendekatan yang banyak di gunakan dalam konteks ini adalah data mining, yang memungkinkan organisasi untuk mengekstrak pola tersembunyi dari data. Data mining memiliki beragam aplikasi di berbagai bidang, termasuk bisnis, ilmu pengetahuan, kesehatan, keuangan, dan lainnya, membantu organisasi untuk memanfaatkan informasi yang tersimpan dalam data mereka untuk tujuan analisis dan pengambilan keputusan yang lebih baik [7], [8]. dan dalam hal ini yang biasa digunakan adalah algoritma K-Means, K-Means merupakan algoritma clustering sederhana yang bersifat tanpa arahan (unsupervised) [6].

Dari berbagai algoritma clustering yang tersedia, Algoritma ini dikenal luas dalam proses pengelompokan karena sifatnya yang sederhana dan efisien. Bahkan IEEE mengakui algoritma ini sebagai salah satu dari sepuluh algoritma data mining terbaik [7]. Clustering merupakan salah satu pendekatan yang digunakan dalam proses data mining [7] metode ini berkerja dengan menyatukan

data yang memiliki atribut serupa ke dalam satu kelompok, dan memisahkan data yang berbeda ke kelompok lain. Namun demikian, kualitas hasil klasterisasi dari algoritma K-Means sangat bergantung pada beberapa faktor penting seperti jumlah klaster (K), karakteristik dataset, serta tahapan preprocessing data yang dilakukan sebelumnya, Proses dimulai dengan memilih secara acak (K) titik pusat kelompok (centroid) di dalam ruang data, lalu mengelompokkan setiap titik data ke dalam kelompok yang memiliki centroid terdekat[9], [10]. Kemudian, titik pusat setiap kelompok dihitung kembali berdasarkan rata-rata titik data dalam kelompok tersebut, dan proses ini diulangi hingga tidak ada lagi perubahan dalam penempatan titik data ke dalam kelompok atau hingga batasan iterasi yang ditentukan tercapai.

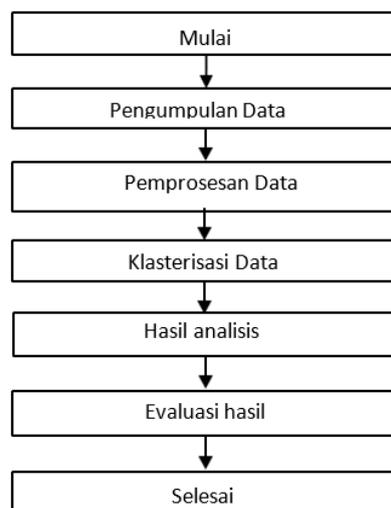
Oleh Karena itu, evaluasi terhadap hasil klasterisasi menjadi langkah krusial dalam proses analisis data. Dua metrik yang umum digunakan untuk menilai kualitas klaster adalah Silhouette Score, yang mengukur sejauh mana data berada dalam klaster yang tepat, dan inertia yang menilai seberapa rapat data dalam satu klaster.

Penelitian ini dilakukan oleh lima peserta yang masing-masing menggunakan dataset berbeda dan parameter yang disesuaikan secara mandiri, termasuk jumlah klaster dan teknik preprocessing. Dengan membandingkan hasil evaluasi klasterisasi dari masing-masing eksperimen, penelitian ini bertujuan memberikan gambaran praktis mengenai bagaimana variasi karakteristik data dan parameter dapat memengaruhi kualitas hasil klasterisasi menggunakan algoritma K-Means.

Berdasarkan latar belakang di atas, maka penelitian ini bertujuan untuk: (1) Untuk Membandingkan Hasil Klasterisasi K-Means Yang Dilakukan Oleh Lima Anggota Kelompok Menggunakan Parameter Dan Dataset Yang Berbeda. (2) Untuk Mengetahui Pengaruh Parameter, Jumlah Klaster, Dan Karakteristik Data Terhadap Nilai Evaluasi Clustering.

METODE

Tahapan dalam penelitian ini dapat dilihat pada Gambar 1 berikut :



Gambar 1. Struktur Penelitian

Pada Struktur Penelitian Diatas Bisa di Jelaskan Pada Penjelasan secara terperinci untuk memahami tahapan demi tahapan yang dilakukan para mahasiswa untuk mendapat hasil evaluasi dari ke 5 dataset.

- 1) **Pengumpulan data:** Pada tahapan pertama yang harus mahasiswa lakukan adalah mengumpulkan data dari berbagai sumber yang ada, para mahasiswa yang beranggotakan 5 orang harus mencari dari buku-buku yang memiliki Sumber data asli yang sah, tetapi ada juga yang menggunakan data contoh yang berasal dari internet agar bisa membandingkan apakah data asli bisa memiliki nilai yang sama atau berbeda dengan data contoh.

- 2) **Pemrosesan data:** Setelah pengumpulan data sudah selesai di lakukan sekarang tinggal menentukan ingin di proses menggunakan algoritma apa yang sesuai dengan data tersebut, tetapi karena sudah memutuskan menggunakan clustering jadi para mahasiswa sepakat menggunakan algoritma K-Means dan juga Sebelum dilakukan klusterisasi, beberapa tahapan preprocessing dilakukan secara mandiri oleh masing-masing anggota sesuai kebutuhan data antara lain normalisasi data numberik, penghapusan data kosong, serta seleksi fitur.
- 3) **Klusterisasi dataset:** karena menggunakan K-Means jadi terlebih dahulu menentukan ingin menggunakan berapa banyak klusterisasi (**K**) dalam 1 dataset, lalu setelah itu di uji menggunakan elbow method dan dilihat apakah inertia nya cocok jika kita memakai klusterisasi begitu banyak.
- 4) **Hasil analisis:** Jika sudah melalui proses untuk menentukan (**K**) dalam dataset yang berbeda sekarang melihat apakah hasil dari klusterisasi sempurna, dengan melihat hasil Silhouette Score pada dataset yang sudah di olah, apabila mendekati 1 atau melewati 0.5 maka klaster bisa dianggap sudah bagus.
- 5) **Evaluasi 5 dataset berbeda:** pada Tahap ini para mahasiswa mulai mencari apakah dataset 1 dengan yang lain nya memiliki kekurangan atau kelebihan dengan menggunakan perbandingan sederhana.

Penelitian ini dilakukan oleh lima mahasiswa yang masing-masing mengolah dataset yang berbeda menggunakan algoritma K-Means untuk melakukan klusterisasi. Dataset yang digunakan meliputi data publik seperti data pertanian, data perbankan, data luas wilayah, data ekspor import dan ada juga data contoh, data berikut ada yang diambil dari internet dan juga ada yang dari data asli yang sudah di catat di masa lalu.

K-Means

K-Means merupakan salah satu metode clustering non-hierarki yang digunakan untuk mempartisi data menjadi dua atau lebih kelompok (cluster). Algoritma ini berkerja dengan membagi data ke dalam klaster, dimana objek-objek dengan karakteristik yang serupa ditempatkan dalam kelompok yang sama , sedangkan objek yang berbeda akan tergabung dalam klaster lainnya [7]. Berkat kesederhanaannya serta kemampuannya dalam menangani data berukuran besar dan keberadaan outlier, K-Means menjadi salah satu algoritma clustering yang paling banyak digunakan. Adapun tahapan dalam pelaksanaan algoritma K-Means menurut[11] adalah sebagai berikut:

1. Menentukan jumlah klaster (K) yang ingin dibentuk.
2. Menentukan titik pusat klaster (centroid) secara acak sebanyak K.
3. Menghitung jarak antara setiap data dan masing-masing centroid menggunakan rumus Euclidean Distance [12] dengan rumus (1) berikut :

$$D(x_i, \pi_i) = \sqrt{\sum_{i=1}^n (x_i - \pi_i)^2} \quad (1)$$

Dimana $d(x_i, \mu_i)$ menyatakan jarak antara data ke- i dan pusat cluster μ , dengan x_i adalah nilai data dan μ_i adalah nilai pusat cluster.

4. Mengelompokkan data ke dalam klaster berdasarkan kedekatan jarak terhadap centroid.
5. Menghitung kembali posisi centroid dengan mengambil rata-rata dari data dalam masing-masing klaster menggunakan persamaan (2):

$$C_k = \frac{1}{n_k} \sum d_i \quad (2)$$

Dimana:

n_k = jumlah data pada klaster ke-K

d_i = jumlah nilai jarak dalam setiap data dalam klaster

6. Mengulang langkah 3 hingga 5 sampai komposisi anggota klaster tidak lagi berubah (konvergen).

Clustering

Clustering atau pengelompokan data adalah proses membagi sekumpulan objek ke dalam beberapa kelompok berdasarkan informasi atau fitur yang tersedia dari data tersebut. Tujuannya adalah untuk memaksimalkan kesamaan antar objek dalam satu kelompok dan meminimalkan kesamaan antar kelompok [6].

Klasterisasi dilakukan menggunakan algoritma K-Means, yang berkerja dengan mencari pusat klaster atau bisa di sebut **centroid**, secara iteratif dan mengelompokan data berdasarkan jarak terdekat ke centroid. Evaluasi terhadap hasil klasterisasi dilakukan dengan dua indikator utama yaitu: **Silhouette Score**, yang mengukur seberapa baik data dengan klaster yang terbentuk, dan Inertia, yang menunjukkan seberapa rapat data dalam satu klaster.

Python

Python adalah bahasa pemrograman dinamis, tingkat tinggi, dimana merupakan bahasa pemrograman interpreter yaitu bahasa yang mengkonversi source code menjadi machine code secara langsung ketika program dijalankan[13]. Semua Proses Analisis Dilakukan Menggunakan Bahasa Pemrograman Python Dengan Bantuan Library Scikit-Learn. Visualisasi Klaster Juga Di Buat Menggunakan Matplotlib Untuk Mempermudah Interpretasi hasil. Berikut adalah potongan kode umum yang digunakan dalam proses klasterisasi:

```
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score

# Menentukan jumlah cluster
k = 3
model = KMeans(n_clusters=k, random_state=42)
model.fit(data)

# Mengambil hasil clustering
labels = model.labels_
inertia = model.inertia_
sil_score = silhouette_score(data, labels)

print("Inertia:", inertia)
print("Silhouette Score:", sil_score)
```

Gambar 2. Potongan kode umum untuk proses Klasterisasi.

HASIL DAN PEMBAHASAN

Berdasarkan teori yang telah dipaparkan di atas, bagian ini akan menjelaskan cara kerja algoritma serta tahapan-tahapan proses yang dilakukan dalam implementasi metode clustering.

Pengumpulan data

Tahap awal dalam pengumpulan data dimulai dengan mengakuisisi data mentah yang dibutuhkan untuk mendukung penelitian ini. Selanjutnya, data tersebut akan melalui tahapan pembersihan (*data cleaning*) guna menghapus informasi yang tidak relevan atau tidak sesuai dengan kebutuhan analisis. Untuk Pengumpulan data, kami memilih untuk mengambil referensi dari sebuah buku dan internet, karena data yang akan kami kumpulkan adalah data asli, yang berarti data yang memang ada atau memiliki bobot pada kehidupan. Untuk beberapa judul buku yang kita ambil adalah seperti ini: Jawa Bandit-Bandit Pedesaan [5]. Pemrograman Python Untuk Penanganan Big Data[3]. Bank Dan Lembaga Keuangan Lain. Bandung[4]. Politik Otonomi Daerah. Profajar Jurnalism[2]. Aplikasi Komputer Dalam Penyusunan Karya Ilmiah[1]. Dari beberapa judul di atas hampir semua judul buku berisikan data set asli.

Pemrosesan data

A. Persiapan data

Pada Bagian Persiapan data kami akan menyajikan dataset dengan berupa tabel seperti di bawah ini.

Tabel 1. Areal Dan Produksi Area Pasuruan dan Probolinggo (1990)

Daerah	Distrik	Perkebunan	Luas (bau)	Produksi/pk
Pasuruan	Pasuruan	Pleret	407	62.593
Pasuruan	Ngampit	Ngampit	620	45.508
Pasuruan	Wangkal	Klurahan	344	550
Pasuruan	Keboncandi	Gayam	524	48.140
Pasuruan	Keboncandi	Pengkol	416	34.545
Pasuruan	Grati	-	-	-
Pasuruan	Winongan	Bekassi Oost	846	85.000
Pasuruan	Rejoso	Sukoharjo	-	-
Bangil	Bangil	Kawisrejo	309	18.158
Bangil	Gempeng	Sumberejo	535	49.416
Bangil	Gempeng	Yosowilangun	-	-
Bangil	Gempeng	Sukorejo	459	47.900
Bangil	Gempol	Ardirejo	606	72.000
Bangil	Gempol	Wangunrejo	388,5	33.600
Malang	Wonoerjo	Arjosari	486,5	36.542
Malang	Gondanglegi	Wonorejo	580	67.415
Malang	Gondanglegi	Sempalwadak	483	43.000
Malang	Gondanglegi	Krebet	-	8.000
Malang	Gondanglegi	Sengoro	-	3.000
Probolinggo	Probolinggo	Umbul	675	63.266
Probolinggo	Dringu	Wonoasih	485	39.971
Probolinggo	Dringu	Wonolangu	475	60.000
Kraksaan	Sumberkarang	Sumberkarang	647	71.500
Kraksaan	Kraksaan	Kandangjati	-	-
Kraksaan	Kraksaan	Seboro	415	42.093
Kraksaan	Kraksaan	Bagu	505	53.000
Paiton	Jalbung	-	462	46.000
Paiton	Pait	-	61.937	77.389
Lumajang	Gending	Gending	-	-
Lumajang	Gending	Maron	596	62.120
Lumajang	Pajarakaan	Pajarakaan	721	84.750
Lumajang	Lumajang	Sukodono	860	83.000
Lumajang	Ranulamongan	Ranupakis	360	30.000

Sumber: Koloniaal Verslag, 1900, lampiran TT.

Data dari Tabel 1 diatas merupakan isi dari buku yang berjudul Jawa Bandit-Bandit Pedesaan.

Tabel 2. Perkembangan Nilai Ekspor dan Impor 1988–1997

Tahun	Termasuk Migas Ekspor	Termasuk Migas Impor	Tidak Termasuk Migas Ekspor	Tidak Termasuk Migas Impor
1988	19.125	13.485	12.945	12.395
1989	21.158	16.356	15.366	13.659
1990	25.759	21.156	19.745	18.454

1991	29.142	23.970	21.802	18.275
1992	29.674	23.987	23.566	18.742
1993	36.822	27.678	26.892	21.376
1994	40.055	29.625	30.578	23.261
1995	45.810	40.286	35.578	26.176
1996	48.144	42.625	36.132	27.314
1997	53.443	41.821	39.512	31.715

Sumber: Statistik Indonesia 1997/BPS

Data dari Tabel 2 diatas merupakan isi dari buku yang berjudul Aplikasi komputer dalam penyusunan karya ilmiah.

Tabel 3. Perkembangan yang Tercatat Di BEI (2011)

Tahun	Kapitalisasi Pasar (Rp triliun)	Perusahaan (unit)
2005	801,25	336
2006	1.249,07	344
2007	1.988,33	383
2008	1.076,49	396
2009	2.019,38	398
2010	3.247,10	420
2011	3.434,45	432

Sumber: CEIC,2011

Data dari Tabel 3 diatas merupakan isi dari buku yang berjudul Bank dan Lembaga Keuangan Lainnya.

Tabel 4. Perbandingan Penduduk dan Luas Wilayah dari Total Nasional

Pulau	Penduduk (%)	Wilayah (%)
Sumatera	21	21
Jawa	57	7
Kalimantan	6	30
Bali-Nusa Tenggara	6	4
Sulawesi	7	8
Papua-Kepulauan Maluku	3	23

Data dari Tabel 2 diatas merupakan isi dari buku yang berjudul Politik otonomi daerah.

Berikut adalah sebuah dataset yang berasal dari 4 buku yang berbeda, tetapi ada 1 data set yang tidak bisa saya tampilkan di sini karena dataset tersebut mempunyai data yang sangat banyak perkiraan bisa mencapai 3000+ data. Oleh karena itu Jika kalian ingin melihat dataset nya di bawah ini kami cantumkan link yang terhubung langsung ke website GitHub.

<https://github.com/mubaris/friendly-fortnight/blob/master/xclara.csv> pada 17 Juni 2025.

B. Pre-Processing

Pada tahapan Pre-Processing, proses pembersihan data (data cleaning) dilakukan menggunakan

perintah sederhana dalam Python. Dalam tahapan ini, variabel atau fitur yang relevan dipilih dari dataset untuk diproses lebih lanjut dalam analisis, dikarenakan seperti pada tabel 1 yang memiliki data kosong dalam kolom tabel yang mana kita harus membersihkan data seperti itu di tiap dataset yang ada. Berikut adalah contoh sederhana dari command Python untuk pembersihan dataset:

```
# 2. Membersihkan data  
df_clean = df.dropna(subset=['Luas (bau)', 'Produksi (pikul)']).copy()  
x = df_clean[['Luas (bau)', 'Produksi (pikul)']].values
```

Gambar 3. Potongan kode umum untuk proses pembersihan data (data cleaning).

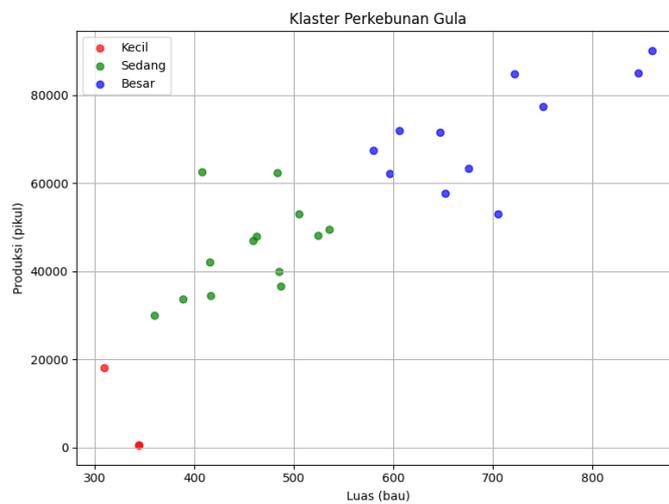
C. Klasterisasi data

Akhir nya tibalah pada tahap klasterisasi yang mana pada setiap mahasiswa untuk mengolah data masing masing agar bisa melihat hasil data akhir yang mana kita mencari Silhouette Score dan inertia dari tiap-tiap dataset.

- **Septiantoro Sogema Fourteen**

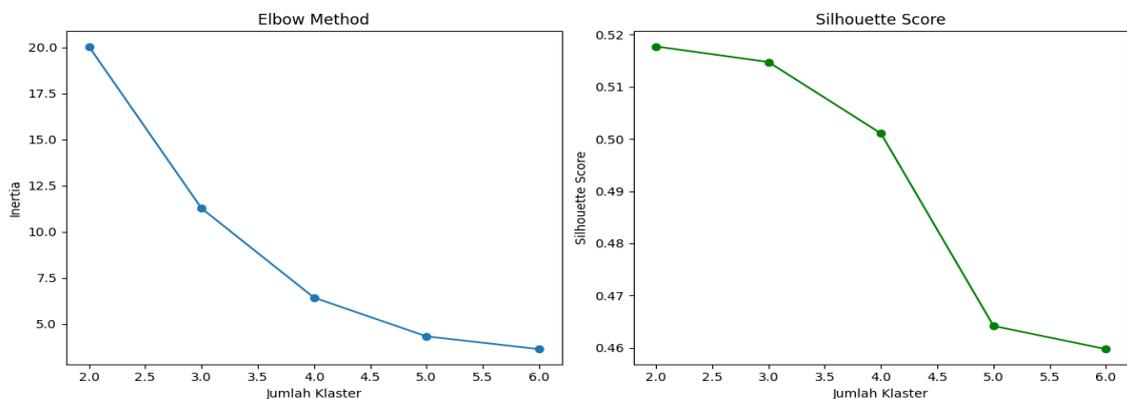
Septiantoro Sogema Fourteen mengolah dataset dari buku yang berjudul Jawa Bandit-Bandit Pedesaan, dengan dataset seperti di tabel nomor 1.

Dan berikut adalah hasil dari klasterisasi untuk dataset dari tabel nomor 1:



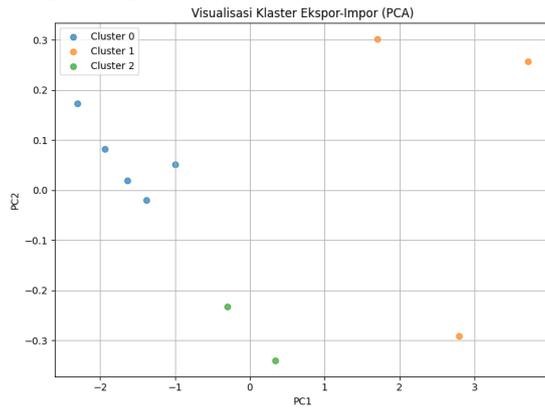
Gambar 4. Hasil Klasterisasi Untuk Dataset Areal dan Produksi Area Pasuruan dan Probolinggo (1990)

Pada gambar untuk hasil klasterisasi di atas menetapkan bahwa $K = 3$, dikarenakan hasil tersebut sudah di perkirakan oleh Septiantoro untuk membuat hanya 3 klastering saja yang terdiri dari, rendah, sedang, dan tinggi nya hasil produksi di tiap tiap daerah. Dan untuk menetapkan bahwa 3 klastering adalah yang terbaik bisa melihat hasil dari Elbow method dan Silhouette Score di bawah ini.

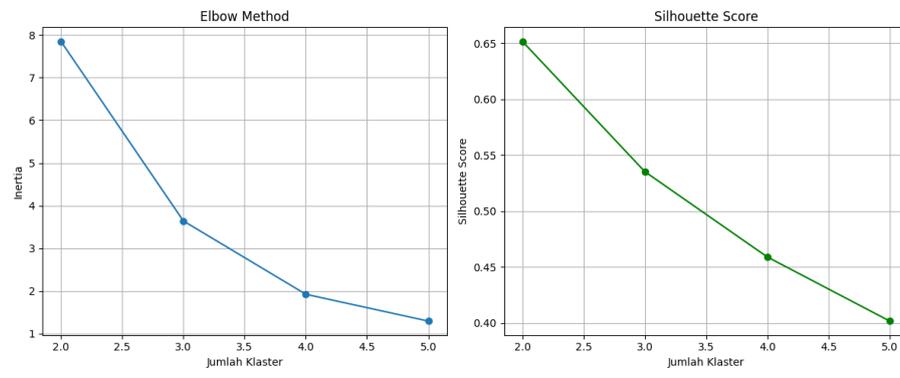


Gambar 5. Hasil klasterisasi yang menampilkan hasil dari Elbow method dan Silhouette Score.

• **AI Hakim Putra Bintang Anugrah**

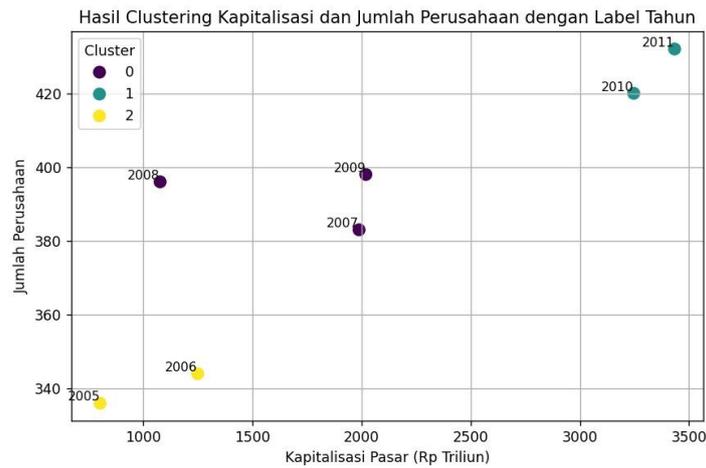


Gambar 6. Hasil Klasterisasi Untuk Dataset Perkembangan Nilai Ekspor dan Impor 1988–1997

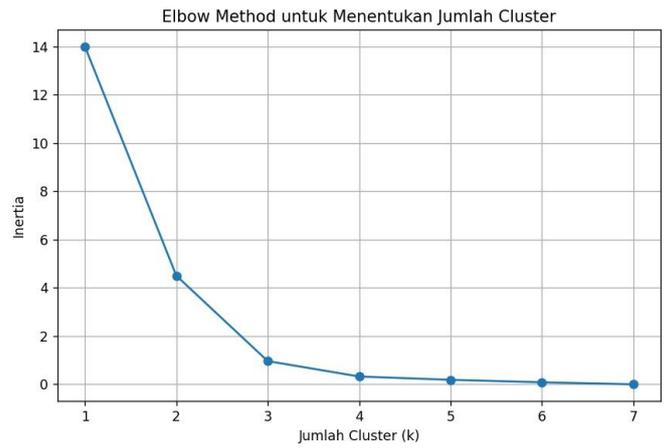


Gambar 7. Hasil Klasterisasi Yang Menampilkan Hasil dari Elbow Method dan Silhouette Score

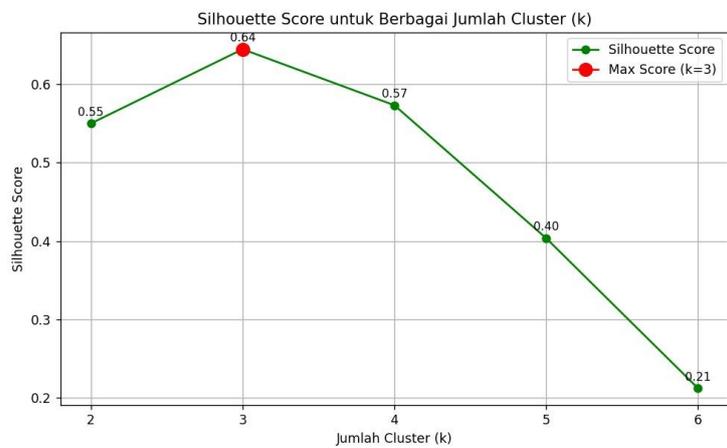
• **Jihan Aulia Shabrina**



Gambar 8. Hasil Klasterisasi Untuk Dataset Perkembangan Yang Tercatat Di BEI (2011)

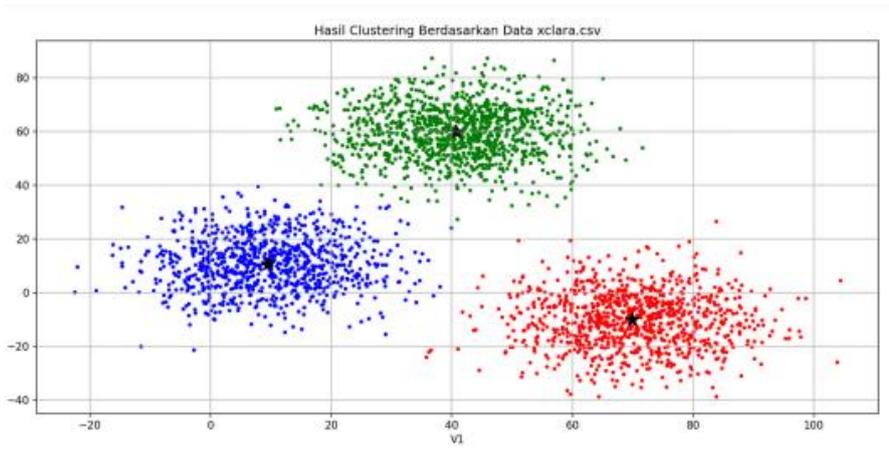


Gambar 9. Hasil klasterisasi yang menampilkan hasil dari Elbow method dan Silhouette Score



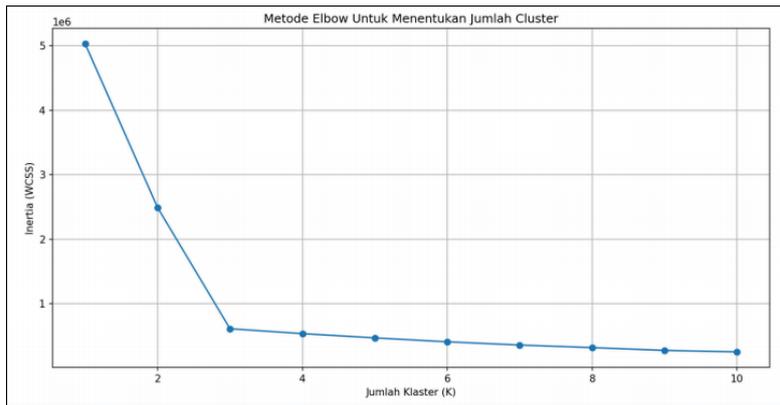
Gambar 10. Hasil Klasterisasi Yang Menampilkan Hasil dari Elbow Method dan Silhouette Score

- **Muhamad Niko Aninda Putra**

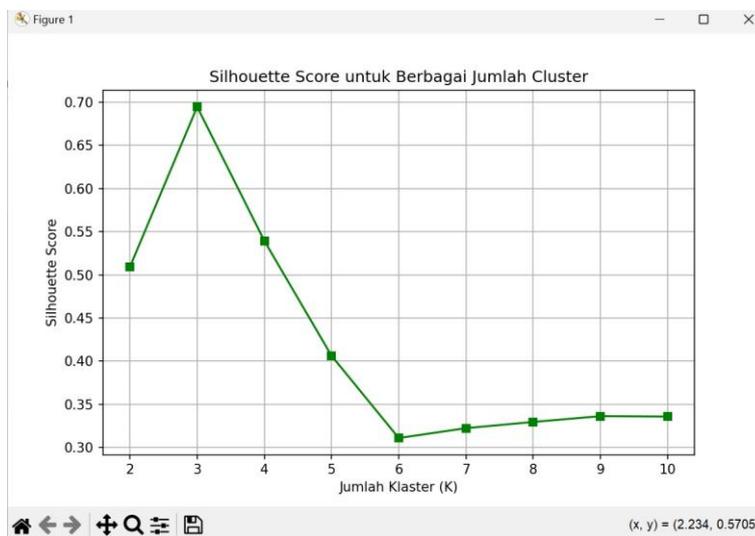


Gambar 11. Hasil Klasterisasi Untuk Dataset

<https://github.com/mubaris/friendly-fortnight/blob/master/xclara.csv> pada 17 Juni 2025.

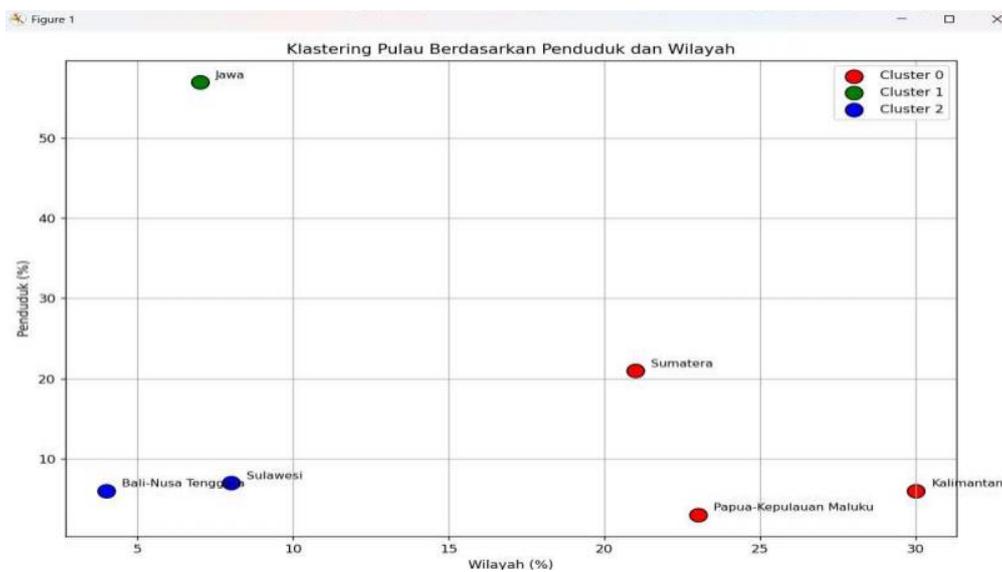


Gambar 12. Hasil Klasterisasi Yang Menampilkan Hasil dari Elbow Method dan Silhouette Score

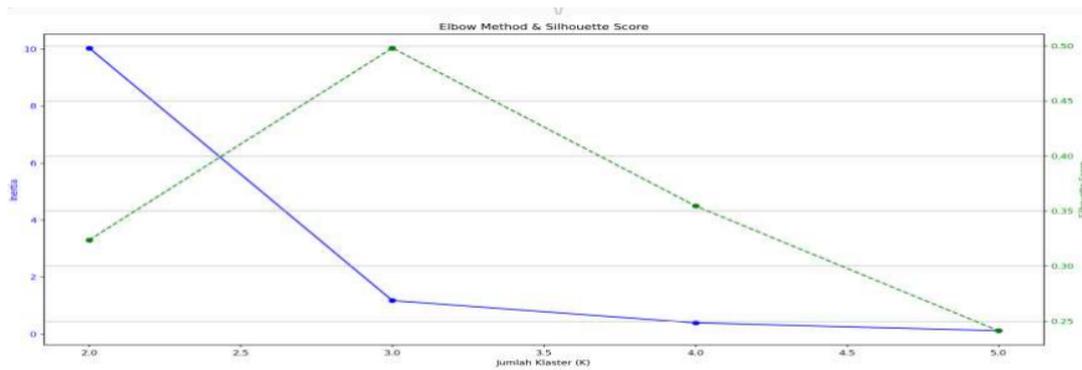


Gambar 13. Hasil Klasterisasi Yang Menampilkan Hasil dari Elbow Method dan Silhouette Score

- Mulyo Adji Winoto



Gambar 14. Hasil Klasterisasi Untuk Datase Perbandingan Luas Wilayah Nasional.



Gambar 15. Hasil Klasterisasi Yang Menampilkan Hasil dari Elbow Method dan Silhouette Score

Pembahasan Perbandingan Antar hasil klasterisasi

Tabel 5. Perbandingan total seluruh hasil klasterisasi

ID	Nama Mahasiswa	Dataset	K	Silhouette Score	Inertia	Catatan
1	Mulyo Aji Winoto	Perbandingan Luas Wilayah Nasional	3	0.50	1.5	Dataset Sederhana Sekala Kecil
2	Jihan Aulia Shabrina	Perkembangan Yang Tercatat Di BEI (2011)	3	0.60	1.5	Perkembangan Saham, Variasi Moderat
3	Septiantoro S.F	Produksi Gula Pasuruan Dan Probolinggo(1900)	3	0.50	11.27	Data Historis, Skala Menengah
4	Muhamad Niko A.P	Dataset GitHub (Mubaris)	3	0.6946	600.000	Dataset Besar, Kemungkinan Banyak Fitur
5	Al Hakim Putra bintang anugrah	Test Of Between-Subject Effects (1950)	3	0.53	3.7	Dataset Eksperimental, Skala Kecil

Data dari Tabel 5 di atas adalah kesimpulan dari seluruh dataset yang telah di proses.

Dari hasil yang diperoleh oleh masing-masing anggota, terlihat adanya variasi yang cukup signifikan dalam nilai **Silhouette Score** dan **Inertia**, meskipun sebagian besar menggunakan jumlah klaster yang sama, yaitu **K =3**.

Mahasiswa dengan nilai silhouette score tertinggi adalah **Muhammad Niko Anindia Putra** dengan nilai **0.6946**, yang menunjukkan bahwa pembagian klaster pada dataset **GitHub** memiliki kualitas pemisahan klaster yang cukup baik, namun demikian, nilai **Inertia** pada data tersebut sangat tinggi (**600.000**), mengindikasikan bahwa jarak antara data terhadap centroid masih sangat

besar, kemungkinan karena **jumlah data yang sangat banyak atau nilai-nilai fitur yang belum di normalisasi**.

Sebaliknya, Mahasiswa Seperti **Jihan Aulia Shabrina** dan **Mulyo adji Winoto** Memiliki Nilai Inertia Rendah (Sekitar 1.2 - 1.5) Dengan Silhouette Score Yang Cukup Baik (0,5-0,6), Menandakan **Dataset Yang Kecil Dan Homogen**. Hal Ini Juga Menunjukkan Bahwa Hasil Clustering Dapat Dianggap Cukup Optimal Pada Konteks Data Skala Kecil.

Sementara Itu, **Septiantoro Sogema Fourteen** Memperoleh Nilai Inertia Sedang (11.27) Dengan Silhouette Score 0.5, Yang Bisa Jadi Dipengaruhi Oleh **Data Historis Yang Tidak Merata Atau Berskala Tidak Seragam**. Sedangkan Al Hakim Putra Mendapatkan Nilai Menengah Untuk Keduanya, Yang Menunjukkan Klasterisasi Cukup Stabil Meskipun Tidak Sangat Optimal.

Secara Umum, Hasil Ini Menunjukkan Bahwa **Hasil Jenis Dataset, Skala Data, Jumlah Fitur, Dan Tahap Preprocessing** Sangat Memengaruhi Kualitas Hasil Klasterisasi. Bahwa Hasil Nilai **K** Disamakan, Perbedaan Karakteristik Data Tetap Menghasilkan Skor Evaluasi Yang Berbeda Secara Signifikan.

SIMPULAN

Jenis dan karakteristik dataset, jumlah fitur, skala data, dan tahapan preprocessing yang dilakukan semua memengaruhi hasil klasterisasi menggunakan algoritma **K-Means**, menurut penelitian ini. Nilai **Silhouette Score dan Inertia** yang dihasilkan menunjukkan perbedaan yang cukup signifikan meskipun seluruh eksperimen menggunakan jumlah klaster yang sama ($K=3$). Agar hasil klasterisasi yang diperoleh lebih bermakna dan optimal, parameter yang tepat dan pemahaman mendalam tentang data diperlukan. Ini karena nilai **Silhouette Score** tertinggi tidak selalu disertai dengan **Inertia** yang rendah, menandakan bahwa ukuran dan kompleksitas data juga memainkan peran penting dalam menentukan kualitas klaster. selain itu, penelitian ini memberikan wawasan praktis bagi pemula yang ingin menilai hasil clustering secara menyeluruh.

DAFTAR PUSTAKA

- Prof.Dr.H.Zaenal Arifin. (2014). Aplikasi Komputer Dalam Penyusunan Karya Ilmiah. Ciledug: Pustaka Mandiri
- Robikin Emhas. (2012). Politik Otonomi Daerah. Profajar Journalism
- Hanna AriniParhusip. (2019). Pemrograman Python Untuk Penanganan Big Data. Yogyakarta: Andi. xclara.csv. Diakses dari <https://github.com/mubaris/friendly-fortnight/blob/master/xclara.csv> pada 17 Juni 2025
- Irham Fahmi, S.E.,M.SI. (2014). Bank Dan Lembaga Keuangan Lain. Bandung: ALFABETA,cv
- Suhartono W. Pranoto. (2010). Jawa Bandit – Bandit Pedesaan. Yogyakarta: Graha Ilmu
- Ramdhan, D., Dwilestari, G., Dinar Dana, R., & Ajiz, A. (2022). Clustering Data Persediaan Barang dengan Menggunakan Metode K-Means. 7(1). <https://media.neliti.com/media/publications/506485-none-96de277a.pdf>
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru. Jurnal Nasional Teknologi Dan Sistem Informasi, 5(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>
- M. S. Sungkar and M. T. Qurohman, "Penerapan Algoritma C5.0 Untuk Prediksi Kelulusan Pembelajaran Mahasiswa Pada Matakuliah Arsitektur Sistem Komputer," J. Media Inform. Budidarma, vol. 5, no. 3, p. 1166, 2021, doi: 10.30865/mib.v5i3.3116.
- M. Cui, "Introduction to the k-means clustering algorithm based on the elbow method," Accounting, Audit. Financ., vol. 1, no. 1, pp. 5–8, 2020.
- A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," J. Artif. Intell. Technol. Inf., vol. 1, no. 1, pp. 20–28, 2023.
- Rahman, A. T. (2017). Coal Trade Data Clusterung Using K-Means (Case Study PT. Global Bangkit Utama).
- Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS. Jurnal Teknologi Dan Sistem Informasi (JTSI), 2(2), 100. <http://jim.teknokrat.ac.id/index.php/sisteminformasi/article/view/868>

Suharto, Agus. *Fundamental Bahasa Pemrograman Python*. Eureka Media Aksara, 2023. [RIS](#)
[\(Mendeley, Zotero, EndNote, RefWorks\) BibTeX \(LaTeX\)](#)