

PENERAPAN ALGORITMA *K-MEANS* UNTUK MENGELOMPOKKAN BERBAGAI JENIS *SMARTPHONE* BERDASARKAN SPESIFIKASI

Gilang Sanjaya¹; Ade Rizki Rinaldi²; Fadhil Muhammad Basysyar³; Fathurrohman⁴;

Program Studi Teknik Informatika¹
Program Studi Rekayasa Perangkat Lunak^{2,4}
Program Studi Sistem Informasi³
STMIK IKMI Cirebon

<https://ikmi.ac.id/page/18/?lang=de>

gillangsanjaya324@gmail.com; aderiziki.ikmi@gmail.com; fadhil.ikmi@gmail.com; fatur.ikmi@gmail.com

(*) Corresponding Author : gillangsanjaya324@gmail.com;

Received: 24 Desember 2024; Accepted: 26 Januari 2025; Published: 31 Desember 2024

Abstract—The increasingly competitive smartphone industry requires companies to better understand consumer preferences and provide products that align with relevant specifications. This study aims to apply the K-Means algorithm to cluster various smartphone types based on technical attributes such as price, RAM, storage capacity, screen size, battery capacity, and camera quality. With this approach, manufacturers and consumers are expected to gain insights to make more accurate product decisions. The research methodology utilizes K-Means clustering algorithm on a dataset comprising 1,021 smartphone data entries retrieved from the Kaggle platform. The data processing involves several stages, including preprocessing, normalization, and clustering analysis, with the optimal number of clusters determined using the Elbow method. The data processing was conducted using the Python programming language, leveraging libraries such as Pandas, NumPy, and Scikit-learn. The clustering process identified three main clusters: the first cluster includes mid-range smartphones suitable for basic needs; the second cluster contains a single smartphone with unique and exclusive specifications; and the third cluster comprises high-end smartphones with advanced specifications, such as large RAM capacities and premium-quality displays. This study contributes to smartphone market segmentation in a more systematic manner. The findings benefit manufacturers in formulating effective marketing strategies and assist consumers in selecting smartphones that meet their needs. For future research, it is recommended to develop more complex recommendation systems based on clustering, incorporating additional machine learning algorithms.

Keywords: K-Means, smartphone, clustering, data analysis, market segmentation

Abstrak—Persaingan yang semakin ketat dalam industri *smartphone* mengharuskan perusahaan untuk lebih memahami keinginan konsumen dan menyediakan produk yang sesuai dengan spesifikasi yang relevan. Penelitian ini bertujuan untuk menerapkan *algoritma K-Means* dalam mengelompokkan berbagai tipe *smartphone* berdasarkan atribut teknis seperti harga, RAM, kapasitas penyimpanan, ukuran layar, daya baterai, dan kualitas kamera. Dengan pendekatan ini, diharapkan produsen dan konsumen dapat memperoleh wawasan dalam mengambil keputusan produk dengan lebih tepat. Metodologi penelitian ini memanfaatkan *algoritma K-Means clustering* menggunakan dataset yang terdiri dari 1.021 data *smartphone* yang diambil dari *platform Kaggle*. Pengolahan data dilakukan melalui beberapa tahapan, termasuk preprocessing, normalisasi, dan analisis *clustering* dengan penentuan jumlah *cluster* yang optimal menggunakan metode *Elbow*. Proses pengolahan data ini dilakukan dengan bahasa pemrograman Python, memanfaatkan pustaka seperti *Pandas*, *NumPy*, dan *Scikit-learn*. Hasil dari proses *clustering* ini mengidentifikasi tiga *cluster* utama. *Cluster* pertama terdiri dari *smartphone* dengan spesifikasi menengah yang cocok untuk kebutuhan dasar. *Cluster* kedua hanya mencakup satu *smartphone* dengan spesifikasi yang unik dan eksklusif. *Cluster* ketiga berisi *smartphone* kelas atas dengan spesifikasi tinggi, seperti kapasitas RAM besar dan layar berkualitas premium. Penelitian ini memberikan kontribusi untuk segmentasi pasar *smartphone* dengan cara yang lebih sistematis. Temuan ini memberikan manfaat bagi produsen dalam merumuskan strategi pemasaran yang efektif dan membantu konsumen untuk memilih *smartphone* yang memenuhi kebutuhan mereka. Untuk penelitian selanjutnya, dianjurkan agar dikembangkan sistem

rekomendasi berbasis *clustering* yang lebih kompleks dengan melibatkan tambahan *algorithm machine learning*.

Kata Kunci : *K-Means, smartphone, clustering*, analisis data, segmentasi pasar

INTRODUCTION

Dunia teknologi semakin kompetitif dan salah satu contohnya adalah *smartphone*. Saat ini berbagai jenis *smartphone* dijual dengan harga dan spesifikasi yang berbeda-beda, sehingga konsumen perlu lebih berhati-hati dalam memilih jenis *smartphone*. Oleh karena itu, perusahaan *smartphone* perlu membedakan dirinya dari merek lain, mengembangkan kebijakan produk melalui pemilihan produk yang mereka hasilkan, dan membuat program penjualan yang mereka tawarkan semenarik mungkin untuk merangsang niat membeli (Almayda & Saepudin, 2021).

Perkembangan teknologi seperti *smartphone* terus meningkat setiap tahunnya di Indonesia. Hal ini tidak lepas dari menjamurnya varian *smartphone* yang berbeda-beda. Salah satu strategi yang bisa Anda gunakan adalah dengan mengelompokkan dan memanfaatkan data penjualan untuk melihat produk mana yang terjual lebih banyak dan produk mana yang terjual lebih sedikit, sehingga Anda dapat mengevaluasi penjualan produk ponsel pintar (Lega et al., 2022).

Banyak penelitian sebelumnya yang mempertimbangkan penggunaan algoritma *K-Means* dalam konteks *Smartphone*. Penelitian ini merujuk pada metodologi sebelumnya, khususnya studi oleh Austin Almayda, Sudin Saepudin, yang menggunakan teknik dan algoritma yang sama untuk mengelompokkan berbagai jenis *smartphone* dengan metode *K-Means clustering* berhasil mengimplementasikan algoritma ini, menghasilkan *cluster* untuk produk *Smartphone* (Almayda & Saepudin, 2021). Selain itu, Penelitian yang ditulis Amri Muliawan Nur, Muhammad Saiful, Hariman Bahtiar, Muhammad Taufik Hidayat, memiliki judul yang hampir sama namun memiliki data yang berbeda. Tujuan dari penelitiannya yaitu mengelompokkan data penjualan *smartphone* yang tepat di Counter Seluler Al Afghani agar lebih mudah, dan berdasarkan hasil pengelompokan x , fokus counter dapat dikurangi untuk menambah persediaan barang yang paling direkomendasikan. (Nur et al., 2024) Terakhir, Penelitian dari Dede Rohman, yang berjudul "*Clustering* Bencana Alam Menggunakan *K-Means* Pada Wilayah Jawa Barat" pada tahun 2024. Tujuan lain dari penelitian ini adalah untuk mengkategorikan insiden bencana alam di Jawa Barat berdasarkan jenis bencananya. Berbagai pendekatan dilakukan, antara lain

mengumpulkan informasi kejadian bencana, mengolah data, menerapkan algoritma *K-Means*, mencari tahu berapa banyak *cluster* yang harus dibuat secara optimal, dan menganalisis temuan *cluster*. Berdasarkan temuan penelitian, Jawa Barat merupakan rumah bagi tiga klaster bencana alam dengan tingkat terjadi yang berbeda-beda: teratas, menengah, dan terbawah (Rohman et al., 2024).

Penelitian ini bertujuan untuk menerapkan algoritma *K-Means* untuk mengelompokkan *smartphone* berdasarkan spesifikasi utama seperti brand, model, RAM, penyimpanan, card, processor, os, ukuran layar dan sebagainya. Dalam penelitian ini, kami ingin menemukan pola pengelompokan yang membantu konsumen memilih *smartphone* yang sesuai dengan kebutuhannya. Penelitian ini juga berupaya untuk mengisi kesenjangan literatur mengenai penerapan algoritma *K-means* dalam pengelompokan data terintegrasi untuk spesifikasi *smartphone*. Selain itu, hasil penelitian ini dapat memberikan wawasan kepada produsen mengenai segmentasi pasar yang lebih tepat sasaran dan mendukung strategi pemasaran yang lebih efektif.

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan algoritma *K-means clustering* untuk mengelompokkan data spesifikasi *smartphone*. Data spesifikasi tersebut diperoleh dari berbagai sumber terpercaya, seperti situs resmi produsen *smartphone* dan *platform* teknologi, serta didukung dengan metodologi serupa dari penelitian sebelumnya yang dilakukan oleh Rahman et al. (2021) Itu telah diambil alih. Setelah data terkumpul, dilakukan tahap prapemrosesan untuk memastikan bahwa data layak untuk dianalisis. Algoritma *K-Means* telah diterapkan untuk mengelompokkan ponsel cerdas berdasarkan atribut tertentu seperti brand, model, RAM, penyimpanan, warna dan harga final seperti dalam karya Suresh et al. (2020) Dia menjelaskan. Jurnal "*Teknik Clustering Data* untuk Segmentasi Pasar". Hasil pengelompokan dievaluasi menggunakan metrik seperti skor siluet untuk mengukur kualitas pengelompokan dan memastikan bahwa hasilnya relevan dengan tujuan penelitian.

Jika penelitian ini mencapai tujuannya, maka hasilnya akan sangat berkontribusi terhadap pemahaman penerapan algoritma *K-means clustering* untuk mengelompokkan data spesifikasi *smartphone*. Hasil penelitian ini diharapkan dapat memperkuat literatur mengenai metode *clustering* di bidang teknologi konsumen, seperti yang

dikemukakan oleh Huang et al.(2021) dia menyarankan. Bahkan, hasil clustering bisa dimanfaatkan konsumen untuk lebih mudah menemukan *smartphone* yang sesuai dengan kebutuhannya. Selain itu, seperti yang dijelaskan Li & Zhang (2021) dalam jurnal *Market Segmentation using Machine Learning Techniques*, produsen ponsel pintar dapat menggunakan wawasan dari pengelompokan ini untuk mengidentifikasi segmen pasar tertentu dan menargetkannya dengan lebih baik. Bagi peneliti, penelitian ini dapat menjadi referensi untuk pertimbangan lebih lanjut penerapan algoritma *K-Means* dalam analisis data spesifik lainnya, seperti perangkat elektronik lainnya dan produk non-teknologi. Di sisi lain, hasil penelitian ini dapat membuka peluang pengembangan sistem rekomendasi berbasis clustering yang lebih kompleks dan relevan, termasuk integrasi dengan model *machine learning* lainnya. Dengan demikian, penelitian ini tidak hanya memberikan kegunaan teoretis tetapi juga nilai praktis yang dapat mendukung pengembangan teknologi dan inovasi basis data.

MATERIALS AND METHODS

Data yang digunakan dalam penelitian ini adalah sekunder yang berasal dari *website kaggle*. Total 1.021 data termasuk 10 atribut: model, harga, baterai, RAM, ukuran layar, penyimpanan, rating, processor, kamera dan OS. Untuk memperoleh data yang sesuai dengan tujuan penelitian, peneliti menggunakan berbagai metode pengumpulan data untuk memastikan bahwa hasil yang diperoleh benar dan dapat digunakan untuk penelitian selanjutnya.(Nur et al., 2024).



Gambar 1. Tahapan Penelitian

Data Pre-Processing

Normalisasi fitur agar setiap atribut berada pada skala yang sama. Hal ini penting karena *K-Means* menggunakan jarak eucliden sehingga perbedaan skala akan sangat berpengaruh dalam hasil *clustering*

Pemilihan Jumlah *cluster*

Elbow Method adalah metode yang tepat untuk mengevaluasi jumlah optimal *cluster* berdasarkan nilai *Sum of squared Errors* (SSE). Pada grafik *Elbow*, titik dimana grafik mulai datar atau membentuk “siku” menandakan jumlah *cluster* yang ideal.

Algoritma *K-Means*

Setiap data di kelompokkan ke *cluster* berdasarkan jarak terdekat dengan *centroid*.

Evaluasi Hasil *Clustering*

Setelah *cluster* terbentuk, dilakukan analisis untuk mengevaluasi hasil *clustering*, apakah *cluster* yang terbentuk memang relevan dengan pola yang diharapkan.

Sumber data ini dari forum *Kaggle*. Forum *Kaggle* ini berisi semua kumpulan data publik. *Kaggle* merupakan *platform* (variabel) penyelenggaraan kontes data ilmiah untuk masalah bisnis, rekrutmen, dan keperluan penelitian akademis, dan data yang digunakan adalah milik Nishantdewal1810 sebagai penulis dan kolaborator. *Dataset* tersebut digunakan dalam penelitian ini mengandung beberapa atribut kunci yang umumnya dianggap penting dalam memilih *smartphone*.(Firmansyah et al., 2023).

- Harga : Mencakup rentang harga berbagai jenis *smartphone*, dari yang murah sampai yang mahal.
- Penyimpanan memori : Besaran memori yang dimiliki *smartphone* misalnya 32GB, 64GB, 128GB dan 256GB.
- Kapasitas baterai : Diukur dalam mAh, yang menunjukkan daya tahan baterai.
- Ukuran layar : Ukuran dalam inci, yang menentukan kenyamanan pengguna dalam menggunakan *smartphone*.
- Resolusi kamera : Mengacu pada kamera belakang dengan resolusi misalnya 12MP, 48MP atau 108MP.

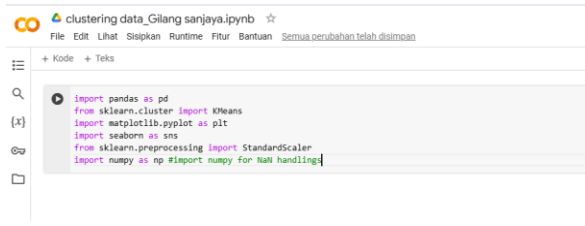
Teknik pengumpulan data pada penelitian ini adalah menggunakan *Algoritma K-means*. Data yang digunakan yaitu data sekunder yang berasal dari *website Kaggle*, dengan jumlah data 1.157 data dengan melibatkan 7 atribut yaitu brand, harga, internal, RAM, kamera, baterai dan *processor*. Untuk mengumpulkan data yang sesuai, peneliti menggunakan teknik pengumpulan data menggunakan *clustering data* dengan *algoritma K-means*

RESULTS AND DISCUSSION

Hasil *Clustering* Menggunakan *K-means*

Penelitian ini menggunakan algoritma *K-Means* untuk mengelompokkan *smartphone* berdasarkan spesifikasi teknisnya seperti rating, RAM, layar, dan spesifikasi lainnya. Dalam hal ini penulis ingin memvisualisasikan hasil dari *clustering data* menggunakan Bahasa *python* pada web *google colab*.

A. Import Library Yang di Butuhkan



Gambar 2 Import Library

Langkah awal yaitu mengimport *library* yang akan digunakan dalam *clustering data* :

1. Pandas as pd digunakan untuk mengolah *dataset*.
2. Sklearn.cluster.Kmeans adalah algoritma yang digunakan yaitu algoritma *K-Means*.
3. Matplotlib dan seaborn digunakan untuk visualisasi data.
4. StandardScaler dari sklearn.preprocessing digunakan untuk menstandarkan data.
5. Numpy digunakan untuk menangani data numerik termasuk NaN (data kosong).

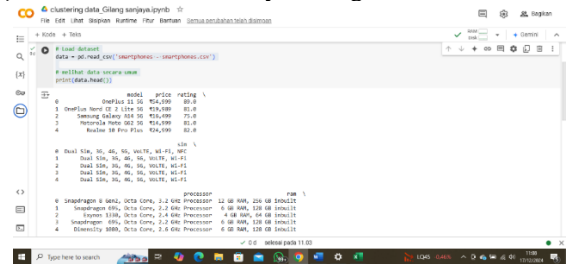
B. Load Dataset



Gambar 3. Load Dataset

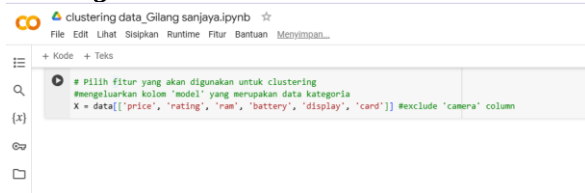
Langkah selanjutnya yaitu *import* data file csv yang berada di samping kiri, data tersebut berupa file yang berisi tentang *smartphone*.

Jika di run maka hasilnya sebagai berikut:



Gambar 4. Hasil Load dataset

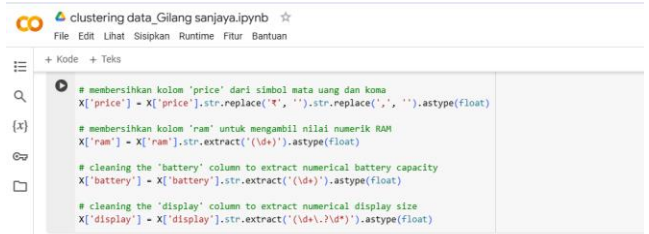
C. Pilih Fitur Yang di Gunakan Untuk Clustering



Gambar 5 Fitur yang digunakan untuk Clustering

Tentukan fitur yang akan digunakan untuk *clustering*, disini penulis memilih fitur *price*, *rating*, *ram*, *battery*, *display*, *card*.

D. Membersihkan Kolom Fitur Numerik

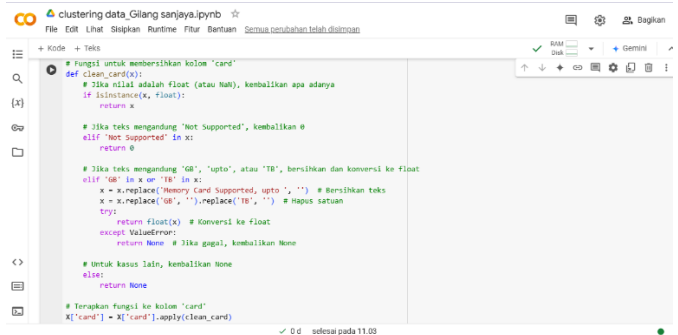


Gambar 6 Membersihkan Kolom Fitur Numerik

Pada bagian ini, kita membersihkan kolom fitur numerik untuk menghapus yang tidak diperlukan untuk *clustering*.

1. Kolom price: membersihkan simbol mata uang (₹) dan koma, lalu di konversi ke dalam tipe float.
2. Kolom ram: mengekstrak angka numerik dari teks agar mendapatkan nilai pada RAM.
3. Kolom battery: membersihkan dan mengambil nilai kapasitas baterai yang formatnya numerik.
4. Kolom display: mengganti ukuran layar dalam format *decimal*.

E. Membersihkan Kolom Card



Gambar 7 Membersihkan Kolom Card

Selanjutnya yaitu membersihkan kolom *card*, fungsi dari *clean_card* yaitu untuk membersihkan kolom *card*.

1. Jika kolom berisi "Not supported" maka nilainya diubah menjadi 0.
2. Jika berisi nilai GB/TB maka teks dibersihkan dan dikonversi ke angka *float*.

Jika terdapat nilai NaN (data kosong) Langkah yang digunakan yaitu menggunakan *.fillna()* untuk mengubah menjadi nilai 0.

```

# Terapkan fungsi ke kolom 'card'
X['card'] = X['card'].apply(clean_card)

# Tangani nilai NaN atau None dengan mengganti ke 0
X['card'] = X['card'].fillna(0)
    
```

Gambar 8 Jika Terdapat Nilai NaN

F. Mengubah Kolom Kategori ke Numerik dengan *One-Hot Encoding*

```

# mengubah fitur kategorial menjadi numerikal menggunakan one-hot encoding
sia_dumales = pd.get_dummies(data['sia'], prefix='sia')
x = pd.concat([sia_dumales, x], axis=1)

processor_dumales = pd.get_dummies(data['processor'], prefix='processor')
x = pd.concat([processor_dumales, x], axis=1)

os_dumales = pd.get_dummies(data['os'], prefix='os')
x = pd.concat([os_dumales, x], axis=1)

# Mengecek jumlah missing values pada kolom rating
print("Jumlah Missing values di kolom rating:", data['rating'].isna().sum())

# Mengisi nilai kosong dengan rata-rata rating
data['rating'].fillna(data['rating'].mean(), inplace=True)
    
```

Gambar 9 Mengubah Kolom Kategori ke Numerik

Pada langkah ini, kolom kategorial seperti *sim*, *processor* dan *os* diubah menjadi format numerik menggunakan *One-Hot-Encoding*, hasilnya akan ditambahkan kembali ke *dataset X*.

G. Normalisasi Data

```

# melakukan normalisasi data
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Validasi data sebelum K-Means
print("Dimensi data:", X_scaled.shape)
print("Tipe data:", type(X_scaled))

# Tangani missing values
X_scaled = np.nan_to_num(X_scaled)

# Pastikan data berbentuk numpy array
if not isinstance(X_scaled, np.ndarray):
    X_scaled = X_scaled.values
    
```

Gambar 10 Normalisasi Data

Selanjutnya yaitu normalisasi data, pada langkah ini ada beberapa point yakni:

1. *StandardScaler* digunakan untuk menstandarkan semua fitur sehingga mempunyai skala yang sama.
2. *X_scaled* berisi data yang di normalisasi.
3. *Np.nan_to_num* untuk menangani nilai NaN (data kosong) dengan diubah menjadi 0.
4. *Np.ndarray* digunakan untuk memastikan data berbentuk *numpy array*

H. Menentukan Jumlah Cluster Dengan *Elbow Method*

```

# Elbow Method
sse = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(X_scaled)
    sse.append(kmeans.inertia_)

plt.plot(range(1, 11), sse, 'o-')
plt.xlabel('Jumlah Cluster (K)')
plt.ylabel('SSE (Sum of Squared Errors)')
plt.title('Elbow Method for Optimal K')
plt.grid(True)
plt.show()

# Check SSE values
for k, error in zip(range(1, 11), sse):
    print(f'K={k}, SSE={error}')
    
```

Gambar 11. Menentukan Jumlah Cluster Dengan *Elbow Method*

Pada penelitian ini metode yang digunakan yaitu metode *Elbow*. Metode *Elbow* adalah cara yang diterapkan untuk menentukan jumlah *cluster* (*c*) yang paling ideal dengan menghitung nilai *Sum of Square Error* (SSE) untuk tiap *cluster*. Semakin besar perbedaan nilai SSE dengan *cluster* berikutnya, yang membentuk titik sudut siku, semakin optimal jumlah *cluster* tersebut. Nilai *c* dalam kombinasi *elbow* dan *K-Means* merupakan representasi grafis dari hubungan antara *cluster* dan pengurangan *error*. Peningkatan nilai *c* menyebabkan grafik secara bertahap menurun hingga nilai *k* mencapai kestabilan. SSE umum digunakan untuk menemukan jumlah *c* yang paling tepat. Teknik ini mengevaluasi nilai SSE dengan berbagai jumlah *cluster* dan kemudian mencari nilai SSE dengan perbedaan terbesar atau yang membentuk sudut paling tajam pada grafik *elbow* untuk menentukan jumlah *cluster* terbaik. Rumus SSE dapat dilihat pada persamaan (Riani et al., 2023).

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|x_i - c_k\|_2^2$$

Pada metode ini meliputi:

1. *K.range* untuk rentang jumlah *cluster* (*K*) yang diuji, dari 1 sampai 10.
2. *Kmeans* untuk algoritma *K-Means* dijalankan untuk setiap nilai *K*.
3. *Inertia_* untuk menunjukkan nilai SSE (*Sum of Squared Errors*), yaitu jumlah kuadrat jarak titik data ke *centroid* terdekat.

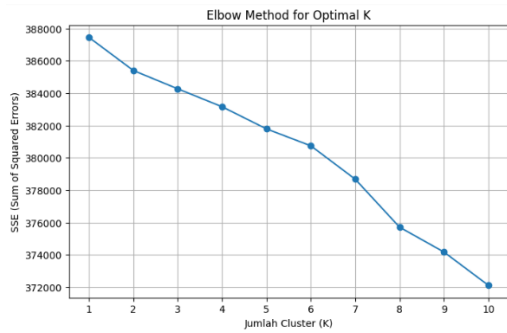
I. Visualisasi *Elbow Method*

```

# Visualisasi Elbow Method
plt.figure(figsize=(8, 5))
plt.plot(range(1, 11), sse, 'o-')
plt.xticks(range(1, 11))
plt.xlabel('Jumlah Cluster (K)')
plt.ylabel('SSE (Sum of Squared Errors)')
plt.title('Elbow Method for Optimal K')
plt.grid(True)
plt.show()

# Check SSE values
for k, error in zip(range(1, 11), sse):
    print(f'K={k}, SSE={error}')
    
```

Gambar 12. Visualisasi *Elbow Method*



Gambar 13. Hasil Visualisasi *Elbow Method*

Pada grafik tersebut menunjukkan hubungan antara jumlah *cluster* (sumbu X) dan nilai *SSE* (sumbu Y). di awal *SSE* turun tajam karena pembagian *data* menjadi lebih banyak *cluster*, setelah beberapa titik penurunan *SSE* menjadi tidak signifikan yaitu membentuk seperti siku.

Berdasarkan grafik, penurunan *SSE* paling tajam terjadi antara **K=1** dan **K=3**. Setelah **K=3**, penurunan *SSE* mulai melambat dan menjadi lebih kecil. Oleh karena itu, jumlah *cluster optimal* berada pada titik **K=3**, karena pada titik tersebut ada keseimbangan antara jumlah *cluster* dan kualitas pengelompokan

J. *K-Means Clustering* dengan Jumlah *Cluster Optimal*

```
clustering_data_Gilang sanjaya.ipynb
File Edit Lihat Sisipkan Runtime Fitur Bantuan Semua perubahan telah disimpan

+ Kode + Teks

# Mengelompokkan data ke dalam 3 cluster
kmeans = KMeans(n_clusters=3, random_state=42)
data['cluster'] = kmeans.fit_predict(X_scaled)

# Menampilkan hasil rata-rata per cluster
numeric_columns = data.select_dtypes(include=np.number).columns
print(data.groupby('cluster')[numeric_columns].mean())
```

Gambar 14 Mengelompokkan data ke 3 *cluster*

Rata-Rata Nilai Kolom Numerik per Cluster:

cluster	rating	cluster
0	78.191771	0.0
1	78.258248	1.0
2	82.176471	2.0

Gambar 15 Hasil Ketika di Run

Setelah menentukan jumlah *cluster* optimal, setelah itu gunakan *K-Means* untuk mengelompokkan data ke dalam 3 *cluster*.

1. **Fit_predict** untuk menjalankan algoritma *K-Means* dan memberi label *cluster* ke setiap baris data.
2. Hasil label *cluster* disimpan dalam kolom baru *cluster* di *dataset* asli.

```
clustering_data_Gilang sanjaya.ipynb
File Edit Lihat Sisipkan Runtime Fitur Bantuan Semua perubahan telah disimpan

+ Kode + Teks

# ===== 5. Menentukan Karakteristik Setiap Cluster =====
# Menampilkan rata-rata kolom numerik untuk setiap cluster
print("\nRata-Rata Nilai Kolom Numerik per Cluster:")
cluster_summary = data.groupby('cluster')[numeric_columns].mean()
print(cluster_summary)

# ===== 6. Analisis Karakteristik Setiap Cluster =====
# Menggunakan analisis statistik untuk melihat distribusi di setiap cluster
for i in range(3):
    print(f"\nKarakteristik Cluster {i}:")
    cluster_data = data[data['cluster'] == i]
    print(f"Jumlah Data: {len(cluster_data)}")
    print(f"Rata-rata Nilai Rating: {cluster_data['rating'].mean()}")
    print(f"Rata-rata Spesifikasi (misalnya RAM, layar, dll):")
    print(cluster_data[numeric_columns].mean())
    print(f"Rata-rata dan variansi untuk setiap fitur:")
    print(cluster_data[numeric_columns].describe())
```

Gambar 15 Menentukan Karakteristik *Cluster*

Setelah data dikelompokkan, langkah selanjutnya yaitu menganalisis karakteristik setiap *cluster* ini dilakukan dengan menghitung rata – rata nilai fitur numerik pada setiap *cluster*.

1. **Groupby('cluster')** untuk mengelompokkan data berdasarkan kolom *cluster*.
2. **Mean()** digunakan untuk menghitung rata – rata setiap kolom numerik untuk masing – masing *cluster*.
3. Hasil ini dapat kita pahami tentang fitur apa yang mendominasi setiap *cluster* dan bagaimana karakteristik *smartphone* pada masing – masing *cluster*.

K. Karakteristik Setiap *Cluster*

Setelah proses *clustering* selesai, hasilnya memperlihatkan bahwa *dataset* dibagi menjadi 3 *cluster* yg berbeda. Berikut ini adalah karakteristik dari 3 *cluster* tersebut.

```
clustering_data_Gilang sanjaya.ipynb
File Edit Lihat Sisipkan Runtime Fitur Bantuan Semua perubahan telah disimpan

+ Kode + Teks

Karakteristik Cluster 0:
Jumlah Data: 1002
Rata-rata Nilai Rating: 78.1917717891634
Rata-rata Spesifikasi (misalnya RAM, layar, dll):
rating: 78.191771
cluster: 0.000000
dtype: float64
Rata-rata dan variansi untuk setiap fitur:
rating cluster
count 1002.000000 1002.0
mean 78.191771 0.0
std 6.265597 0.0
min 68.000000 0.0
25% 75.000000 0.0
50% 78.258248 0.0
75% 83.000000 0.0
max 89.000000 0.0

Karakteristik Cluster 1:
Jumlah Data: 1
Rata-rata Nilai Rating: 78.25824800910125
Rata-rata Spesifikasi (misalnya RAM, layar, dll):
rating: 78.258248
cluster: 1.000000
dtype: float64
Rata-rata dan variansi untuk setiap fitur:
rating cluster
```

Gambar 16 Hasil Karakteristik *Cluster* Ketika di Run

```
clustering_data_Gilang sanjaya.ipynb
File Edit Lihat Sisipkan Runtime Fitur Bantuan Semua perubahan telah disimpan

+ Kode + Teks

rating cluster
count 17.000000 17.0
mean 82.176471 1.0
std 2.617825 0.0
min 79.000000 0.0
25% 80.000000 2.0
50% 82.000000 2.0
75% 84.000000 2.0
max 87.000000 2.0
```

Gambar 16 Hasil Karakteristik *Cluster* Ketika di Run

a. Cluster 0

1. Jumlah *Data*: 1002 *smartphone* tergolong dalam kelompok ini.
2. Rata-rata Nilai *Rating*: Rata-rata penilaian untuk *smartphone* di kelompok ini adalah 78.19. Nilai penilaian ini cukup tinggi dan relatif stabil dalam kelompok ini.
3. Rata-rata Spesifikasi:

Penilaian memiliki rata-rata **78.19**, menunjukkan bahwa *smartphone* dalam kelompok ini umumnya memiliki penilaian yang baik.

Variansi untuk kolom kelompok adalah 0, yang menunjukkan semua data dalam kelompok ini tergabung pada kelompok yang sama (kelompok 0).

- a. *Mean* (Rata-rata): 78.19
- b. *Min*: 60 (penilaian terendah)
- c. *Max*: 89 (penilaian tertinggi)
- d. Standar Deviasi (std): 6.91, menunjukkan variasi yang cukup luas dalam penilaian, tetapi tetap berada dalam kisaran yang tidak terlalu lebar.

b. Cluster 1

1. Jumlah *Data*: 1 *smartphone* terdaftar dalam kelompok ini.
2. Rata-rata Nilai *Rating*: Nilai penilaian untuk *smartphone* ini adalah **78.26**, yang sangat mirip dengan kelompok 0.
3. Rata-rata Spesifikasi:
 - a. *Rating* 78.26.
- b. Kelompok untuk *smartphone* ini adalah 1 (termasuk dalam kelompok 1).
- c. Statistik Deskriptif: Karena hanya ada satu data, tidak ada variasi (standar deviasi dan variansi tidak ada), dan nilai minimum, median, serta maksimum semuanya adalah 78.26.

c. Cluster 2

1. Jumlah *Data*: 17 *smartphone* yang tergabung dalam kelompok ini.
2. Rata-rata Nilai *Rating*: Rata-rata penilaian di kelompok ini adalah **82.18**, lebih tinggi dibandingkan dengan kelompok 0 dan 1.
3. Rata-rata Spesifikasi:
 - a. Penilaian untuk *smartphone* dalam kelompok ini adalah 82.18, yang menunjukkan kualitas penilaian yang sedikit lebih tinggi.
 - b. Variansi untuk kolom kelompok adalah 0, yang berarti semua data dalam kelompok ini tergabung pada kelompok yang sama (kelompok 2).
4. Statistik Deskriptif:
 - a. *Mean* (Rata-rata): 82.18
 - b. *Min*: 78 (penilaian terendah)
 - c. *Max*: 87 (penilaian tertinggi)
 - d. Standar Deviasi (std): 2.63, menunjukkan variasi yang lebih rendah dibandingkan

dengan kelompok 0, sehingga penilaian di kelompok ini lebih konsisten.

Ringkasan dari penjelasan karakteristik *cluster* sebagai berikut :

1. Kelompok 0 berisi mayoritas *data* dengan penilaian rata-rata 78.19, menunjukkan kualitas penilaian yang cukup baik, namun dengan variasi yang lebih besar.
2. Kelompok 1 berisi hanya satu *data*, dengan penilaian 78.26, sangat mirip dengan kelompok 0, tetapi tidak memberikan wawasan lebih lanjut karena hanya ada satu data.
3. Kelompok 2 berisi sedikit *data*, tetapi penilaiannya sedikit lebih tinggi (rata-rata 82.18) dan lebih konsisten dibandingkan dengan kelompok lainnya, dengan variasi penilaian yang lebih rendah.

CONCLUSION

Studi ini berhasil mengimplementasikan algoritma *K-Means* untuk mengelompokkan berbagai model *smartphone* berdasarkan spesifikasi teknisnya, seperti RAM, kapasitas penyimpanan, ukuran layar, daya baterai, harga, dan kualitas kamera. Melalui analisis *dataset* yang mencakup 1.021 data *smartphone*, tiga kelompok utama berhasil diidentifikasi:

1. *Cluster 0: Smartphone premium* dengan spesifikasi tinggi, seperti RAM besar dan layar berkualitas tinggi.

2. *Cluster 1: Sebuah unit smartphone* dengan spesifikasi eksklusif dan unik yang hanya digunakan untuk pengguna yang membutuhkan fitur dan esensial.

3. *Cluster 2: Smartphone* dengan spesifikasi menengah (*mid-range*) yang ideal untuk kebutuhan dasar.

REFERENCE

- Ainur Rahman, & Suroyo, H. (2021). Analisis Data Produk Elektronik Di E-Commerce Dengan Metode Algoritma K-Means Menggunakan Python. *Journal of Advances in Information and Industrial Technology*, 3(2), 11-18. <https://doi.org/10.52435/jaiit.v3i2.158>
- Almayda, A., & Saepudin, S. (2021). PENERAPAN DATA MINING K-MEANS CLUSTERING UNTUK MENGELOMPOKKAN BERBAGAI JENIS MERK SMARTPHONE.
- Azhar, Z., Wulandari, C., Hanum, Z., Putra, W. A., & Saragih, Y. P. (2024). Implementasi Pengelompokan Persediaan Sepeda Motor Menggunakan Metode Clustering K-Means. In

- Journal of Computer Science and Information Technology E-ISSN* (Vol. 4, Issue 2).
- Firmansyah, M. I., Rohman, R. S., Marsusanti, E., Bina, U., Informatika, S., & Disetujui, D. D. (2023). Penerapan Algoritma Klastering K-Means Untuk Fitur Atribut Pada Layanan Streaming Musik Spotify. *Journal Computer Science*, 2(2).
- Gulo, A., & Anis, M. (2024). *Biner: Jurnal Ilmu Komputer, Teknik dan Multimedia Analisa Dan Perancangan Sistem Pengelompokan Data Rekam Medis Pasien Pada Puskesmas Cipayung Depok Menggunakan Metode K-Means Berbasis Web*. 2(1). <https://journal.mediapublikasi.id/index.php/Biner>
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>
- Handoko, S., Fauziah, F., & Handayani, E. T. E. (2020). IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 25(1), 76–88. <https://doi.org/10.35760/tr.2020.v25i1.2677>
- Hasibuan, F. P. A., Sumarno, S., & Parlina, I. (2021). Penerapan K-Means pada Pengelompokan Penjualan Produk Smartphone. *SATESI: Jurnal Sains Teknologi Dan Sistem Informasi*, 1(1), 15–20. <https://doi.org/10.54259/satesi.v1i1.3>
- Khan, A. S. S., Fatekurohman, M., & Dewi, Y. S. (2023). Perbandingan Algoritma K-Medoids Dan K-Means Dalam Pengelompokan Kecamatan Berdasarkan Produksi Padi Dan Palawija Di Jember. *Jurnal Statistika Dan Komputasi*, 2(2), 67–75. <https://doi.org/10.32665/statkom.v2i2.2301>
- Lega, A., Adytia, P., Lailiyah, S., Studi, P., Informatika, T., Widya, S., Dharma, C., Yamin, J. M., & 25, N. (n.d.). *PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK KLASTERISASI PENJUALAN SMARTPHONE PADA CARIN CELL*.
- Marisa, F., Zahma, A., Mui Bau, A., Noviansa, E., Neno, A. S., Lidya Maukar, A., Informatika, T., Malang, W., Borobudur, J., & 35, N. (2021). Digitasi Produktivitas Panen Padi Berbasis K-Means Clustering. *SMARTICS Journal*, 7(1). <https://doi.org/10.21067/smartics.v7i1.5270>
- Nur, A. muliawan, Saiful2, M., Bahtiar, H., & Muhammad Taufik Hidayat. (2024). Penerapan Algoritma K-Means Clustering Dalam Mengelompokkan Smartphone Yang Rekomendasi Berdasarkan Spesifikasi. *Infotek: Jurnal Informatika Dan Teknologi*, 7(2), 478–488. <https://doi.org/10.29408/jit.v7i2.26283>
- Oktaviani, S., & Bahtiar, A. (2023). Implementasi Algoritma K-Means Dalam Pengelompokan Data Penjualan CV. Widuri Menggunakan Orange. In *Jurnal Wahana Informatika (JWI)* (Vol. 2, Issue 1).
- Priyatman, H., Sajid, F., & Haldivany, D. (2019). *JEPIN (Jurnal Edukasi dan Penelitian Informatika) Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa*.
- Putri Riyandoro, A., Voutama, A., Umaidah, Y., Ronggo Waluyo, J. H., Timur Karawang, T., & Barat, J. (2023). IMPLEMENTASI DATA MINING CLUSTERING K-MEANS DALAM MENGGOLONGKAN BERAGAM MEREK LAPTOP. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 2).
- Riani, A. P., Voutama, A., & Ridwan, T. (2023). Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow. *J-SISKO TECH (Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD)*, 6(1), 164. <https://doi.org/10.53513/jsk.v6i1.7351>
- Rizki Risdah Sitorus, & Rafiqi Aidil Fitra. (2023). Penerapan Metode K-Means Clustering pada Hasil Produksi Beras di Wilayah Sumatera Utara. *Mutiara: Jurnal Penelitian Dan Karya Ilmiah*, 1(6), 375–381. <https://doi.org/10.59059/mutiara.v1i6.749>
- Rohman, D., Annisa, R., Efendi, D. I., & Solahudin, D. (2024). CLUSTERING BENCANA ALAM MENGGUNAKAN K-MEANS PADA WILAYAH JAWA BARAT. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 1). <https://www.ncbi.nlm.nih.gov/books/NBK55>
- Sugianto, C. A., Rahayu, A. H., & Gusman, A. (2020). Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Cigugur Tengah. *Journal of Information Technology*, 2(2), 39–44. <https://doi.org/10.47292/joint.v2i2.30>
- Sulistiawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*, 15(2), 25. <https://doi.org/10.33365/jtk.v15i2.1162>