

Pengelompokan Smartphone Berdasarkan Spesifikasi Menggunakan Algoritma K-Means untuk Analisis Segmentasi Produk

¹Anis Faizul Laila, ²Putri Ameliya, ³Erna Daniati

¹⁻³ Sistem Informasi, Universitas Nusantara PGRI Kediri

E-mail: ¹anisfaizul879@gmail.com, ²pameliya977@gmail.com, ³ernadaniati@unpkediri.ac.id

Penulis Korespondens : Erna Daniati

Abstrak— Variasi smartphone dengan harga dan spesifikasi yang beragam menuntut analisis karakteristik produk yang tepat. Penelitian ini bertujuan mengelompokkan data smartphone berdasarkan atribut harga, rating, dan RAM menggunakan metode K-Means Clustering. Data diambil dari platform Kaggle dan diproses melalui tahap preprocessing seperti penanganan data kosong dan standarisasi. Metode Elbow digunakan untuk menentukan jumlah kluster optimal, yang menghasilkan tiga kluster utama: kelas atas, menengah, dan bawah. Hasil klasterisasi ini dapat membantu memahami distribusi pasar smartphone dan mendukung strategi pemasaran yang lebih terarah. Penelitian ini memberikan kontribusi dalam segmentasi produk yang informatif bagi produsen dan konsumen.

Kata Kunci—Clustering, K-Means, Segmentasi Produk

Abstract— The variety of smartphones with different prices and specifications requires proper product characteristic analysis. This study aims to group smartphone data based on price, rating, and RAM attributes using the K-Means Clustering method. Data is taken from the Kaggle platform and processed through preprocessing stages such as handling empty data and standardization. The Elbow method is used to determine the optimal number of clusters, which results in three main clusters: high-end, middle-end, and low-end. The results of this clustering can help understand the distribution of the smartphone market and support more targeted marketing strategies. This study contributes to informative product segmentation for manufacturers and consumers.

Keywords—Clustering, K-Means, Product Segmentation

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Kemajuan teknologi komunikasi, khususnya di bidang smartphone, mendorong persaingan yang sangat ketat antar produsen dalam menghadirkan produk dengan spesifikasi canggih dan harga yang kompetitif [1]. Situasi ini menyebabkan pasar smartphone menjadi sangat kompleks, di mana konsumen dihadapkan pada banyak pilihan produk dengan berbagai variasi spesifikasi seperti kapasitas RAM, kamera, dan penyimpanan [2]. Kompleksitas ini berpotensi menyebabkan kebingungan konsumen dalam menentukan pilihan yang tepat sesuai kebutuhan dan anggaran mereka, yang akhirnya dapat berdampak pada keputusan pembelian yang kurang optimal [3].

Beberapa penelitian sebelumnya telah mengangkat isu segmentasi pasar produk smartphone sebagai solusi untuk mengatasi masalah kompleksitas ini. Misalnya, [4] menggunakan teknik clustering untuk mengelompokkan produk elektronik berdasarkan fitur teknisnya guna meningkatkan efektivitas pemasaran. Namun, pendekatan tersebut masih kurang mempertimbangkan faktor harga sebagai variabel utama dalam segmentasi, yang sebenarnya sangat krusial dalam pengambilan keputusan konsumen. Selanjutnya, penelitian oleh [5] menunjukkan bahwa algoritma K-Means efektif dalam mengelompokkan produk berdasarkan data numerik, tetapi aplikasinya masih terbatas pada kategori produk tertentu dan belum mengintegrasikan analisis dampak segmentasi terhadap strategi bisnis secara mendalam. Dengan menerapkan metode ini dalam konteks data mining, pengelolaan dan analisis data spesifikasi smartphone dapat dilakukan secara sistematis, menghasilkan informasi yang relevan untuk pengambilan keputusan [6].

Dengan mempertimbangkan kekurangan tersebut, penelitian ini berupaya memberikan kontribusi baru dengan fokus mengelompokkan smartphone berdasarkan dua variabel utama, yaitu harga dan kapasitas RAM, menggunakan algoritma K-Means Clustering. Pendekatan ini diharapkan dapat menghasilkan segmentasi produk yang lebih terstruktur dan informatif, sekaligus memberikan nilai tambah bagi produsen dalam menyusun strategi pemasaran yang lebih tepat sasaran serta memudahkan konsumen dalam memilih produk yang sesuai preferensi dan anggaran. Penelitian oleh [7] memperkuat kerangka berpikir ini dengan menekankan pentingnya segmentasi pasar berbasis data teknis dan harga untuk meningkatkan daya saing produk di pasar elektronik. Penelitian ini bertujuan menjawab permasalahan utama: bagaimana mengelompokkan produk smartphone secara informatif berdasarkan harga dan RAM untuk mendukung strategi pemasaran yang efektif. Kebaruan penelitian ini terletak pada penggabungan variabel harga dan kapasitas RAM dalam segmentasi produk menggunakan K-Means, yang sebelumnya jarang dikaji secara bersamaan dalam konteks pasar smartphone Indonesia. Dengan demikian, hasil penelitian ini diharapkan memberikan kontribusi signifikan baik dari sisi akademis maupun praktis dalam pengelolaan pemasaran produk teknologi.

II. METODE

Penelitian ini bertujuan untuk menerapkan K-Means Clustering dalam mengidentifikasi kelompok smartphone berdasarkan harga, rating, dan kapasitas RAM. Pengelompokan ini diharapkan dapat membantu konsumen dalam menentukan pilihan dan produsen dalam menyusun strategi pemasaran yang lebih terarah [8]. Analisis dilakukan menggunakan Google Colab dengan

bahasa pemrograman Python, yang memungkinkan implementasi algoritma K-Means secara efisien dan terstruktur.

2.1 Pengumpulan Data

Tahapan pertama dari metode penelitian adalah melakukan pengumpulan data. Penelitian ini menggunakan dataset publik dari Kaggle, yaitu Smartphone Models Dataset. Dataset ini merupakan kumpulan data Ponsel Pintar, (Gambar 1) menunjukkan field yang digunakan berisi informasi tentang merek, model, harga, dan kapasitas RAM smartphone, dengan total 785 data setelah melalui proses pra-pemrosesan.

	brand_name	model	price	rating	ram_capacity
0	oneplus	OnePlus 11 5G	54999	89.0	12.0
1	oneplus	OnePlus Nord CE 2 Lite 5G	19989	81.0	6.0
2	samsung	Samsung Galaxy A14 5G	16499	75.0	4.0
3	motorola	Motorola Moto G62 5G	14999	81.0	6.0
4	realme	Realme 10 Pro Plus	24999	82.0	6.0

Gambar 1: Dataset

2.2 Data Preprocessing

Salah satu metode yang umum digunakan dalam data mining untuk mengidentifikasi serta menemukan pola dalam kumpulan data adalah *Knowledge Discovery in Databases* (KDD). Proses ini terdiri dari beberapa langkah utama, yaitu: pemilihan data, pembersihan awal, transformasi, eksplorasi data menggunakan teknik data mining, serta evaluasi atau interpretasi hasil [9]. Tahap awal, yaitu pemilihan data, berfokus pada pengambilan data yang sesuai dengan tujuan penelitian. Selanjutnya, pada tahap pembersihan awal atau pra-pemrosesan, dilakukan penyaringan terhadap data yang tidak konsisten, duplikat, tidak relevan, atau mengandung gangguan (noise). Proses kemudian dilanjutkan dengan transformasi data, yakni mengubah format data agar sesuai dengan kebutuhan analisis. Langkah utama dari KDD adalah penerapan teknik data mining untuk mengekstraksi pola-pola bermakna dari data. Setelah itu, hasil yang diperoleh dievaluasi untuk menilai keakuratan serta relevansi informasi yang dihasilkan terhadap permasalahan yang diteliti oleh [10] penelitian ini, proses pembersihan meliputi penghapusan data yang tidak berkaitan dan duplikat, serta mengkonversi nilai harga ke dalam satuan rupiah agar selaras dengan konteks pasar di Indonesia.

2.3 Metode Clustering dengan K-Means

Algoritma K-Means dimanfaatkan untuk mengelompokkan perangkat berdasarkan kesamaan spesifikasi sebagai tahap awal sebelum dilakukan analisis lanjutan atau penyusunan strategi rekomendasi produk [11]. Metode ini bertujuan untuk membentuk kelompok data berdasarkan tingkat kemiripan karakteristik, seperti harga dan kapasitas RAM, yang dimiliki oleh setiap smartphone [12]. Setiap perangkat akan diklasifikasikan ke dalam satu cluster tertentu sesuai dengan kemiripan atribut yang dimiliki dibandingkan perangkat lainnya. Proses ini mengandalkan perhitungan jarak antara masing-masing data dan pusat cluster (centroid) untuk menilai tingkat kemiripannya [13]. Jarak yang digunakan dalam algoritma ini umumnya dihitung dengan rumus *Euclidean Distance*, yang memungkinkan pemetaan data secara

akurat dalam ruang multidimensi [14]. Penggunaan K-Means Clustering pada data smartphone sangat berguna karena mampu memproses dan membagi data dalam skala besar secara cepat dan efisien.

Proses pada algoritma K-Means terdiri dari beberapa tahapan perhitungan yang harus dijalankan, antara lain:

- a. Tetapkan jumlah clustering yang diinginkan (k).
- b. Menentukan centroid untuk setiap kelompok, di mana centroid berfungsi sebagai titik pusat yang mewakili karakteristik keseluruhan dari kelompok tersebut, layaknya seorang “pemimpin”.
- c. Hitung jarak antara setiap data dengan centroid yang ada, kemudian tempatkan data ke dalam kelompok yang memiliki centroid paling dekat.
- d. Tentukan centroid untuk setiap kelompok berdasarkan data yang telah tergabung, lalu hitung kembali jarak antara data dan centroid yang baru.
- e. Proses sebelumnya diulangi secara terus-menerus hingga susunan anggota dalam setiap kelompok tetap dan tidak mengalami perubahan lagi.

III. HASIL DAN PEMBAHASAN

3.1 Preprocessing dan Eksplorasi Data

3.1.1 Pembersihan Data

Dataset awal berisi berbagai fitur, tetapi hanya kolom `brand_name`, `model`, `price`, `price_idr`, `rating`, dan `ram_capacity` yang digunakan untuk analisis karena relevansinya dengan segmentasi pasar. Nilai yang hilang pada kolom `rating` dihapus, dan dataset difilter untuk menghilangkan outlier, menghasilkan 785 catatan.

	price	rating	ram_capacity
count	878.000000	878.000000	878.000000
mean	29701.361048	78.249431	6.552392
std	37356.823164	7.402454	3.141855
min	3499.000000	60.000000	2.000000
25%	13499.000000	74.000000	4.000000
50%	19990.000000	80.000000	6.000000
75%	31994.000000	84.000000	8.000000
max	650000.000000	89.000000	64.000000

Gambar 2 : Statistik Deskriptif Fitur yang Dipilih

3.1.2 Data Transformation

Konversi harga dari Rupee India (INR) ke Rupiah Indonesia (IDR) dilakukan dengan menggunakan nilai tukar rata-rata sebesar $1 \text{ INR} \approx 192,06 \text{ IDR}$. Kode untuk konversi harga terdapat pada (Gambar 3)

```
# Konversi INR ke IDR
exchange_rate = 192.06
df['price_idr'] = (df['price'] * exchange_rate).round(0)

# Tampilkan kolom harga awal dan harga setelah dikonversi
print(df[['price', 'price_idr']].head())
exchange_rate = 192.04
df['price_idr'] = df['price'] * exchange_rate
df[['price', 'price_idr']].head()
```

Gambar 3 : Konversi Harga dari INR ke IDR

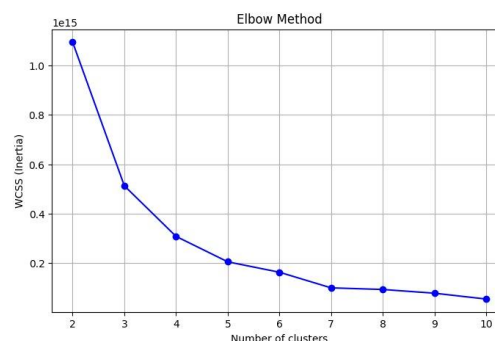
3.2 Metode Clustering

3.2.1 Standarisasi Fitur

Fitur `price_idr`, `rating`, dan `ram_capacity` distandarisasi menggunakan `StandardScaler` dari `scikit-learn` untuk memastikan bobot yang setara antar fitur. Standarisasi mengubah data sehingga memiliki rata-rata 0 dan standar deviasi 1, mengatasi perbedaan skala (misalnya, harga dalam jutaan IDR vs. RAM dalam GB). Standarisasi dilakukan setelah konversi harga ke IDR untuk konsistensi dengan analisis lokal.

3.2.2 Penentuan Jumlah Kluster Optimal

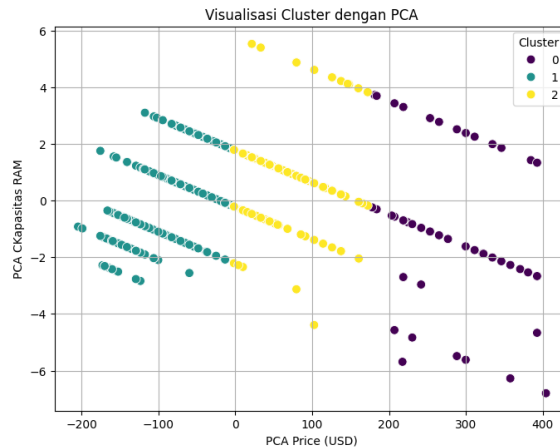
Jumlah kluster optimal (k) ditentukan menggunakan **Metode Elbow**: Within-Cluster Sum of Squares (WCSS) dihitung untuk (k) dari 1 hingga 10. Plot WCSS (Gambar 4) menunjukkan tikungan (k) = 3 hingga (k) = 10, mengindikasikan bahwa penambahan kluster di luar rentang ini menghasilkan pengurangan WCSS yang semakin kecil.



Gambar 4 : Plot Metode Elbow

3.2.3 Klastering K-Means

Proses pengelompokan data dilakukan menggunakan bahasa pemrograman Python dengan bantuan pustaka `scikit-learn`. Parameter yang digunakan pada algoritma K-Means antara lain `n_clusters=3` dan `random_state=0`, sementara pengaturan lainnya mengikuti nilai bawaan dari `scikit-learn`. Berdasarkan [15], `scikit-learn` merupakan pustaka open-source yang cukup lengkap dan banyak digunakan dalam bidang pembelajaran mesin. Pustaka ini mendukung empat aspek utama, yaitu transformasi data, pembelajaran terawasi, pembelajaran tidak terawasi, serta evaluasi dan pemilihan model. Kluster divisualisasikan dalam ruang 2D menggunakan PCA untuk mengurangi dimensi fitur (Gambar 5).



Gambar 5 : Visualisasi PCA Klaster

Dataset diperkaya dengan label klaster, dan rata-rata fitur dihitung untuk setiap klaster berdasarkan hasil yang diberikan (Tabel 1). Smartphone representatif dari setiap klaster dianalisis untuk memahami posisi pasarnya.

Tabel 1 : Ringkasan Cluester

cluster	price_idr	rating	ram_capacity
0	8.419.303,29	84,63	8.464286
1	2.565.787,42	74,06	5.156448
2	5.114.490,95	82,40	7.644737

1. Klaster 0 : Segmen High-end
Klaster ini berisi smartphone dengan harga tinggi, RAM besar, dan rating pengguna yang tinggi.
 - Rata-rata Harga : Rp8.419.303
 - Rata-rata Rating: 84,63
 - Rata-rata RAM: 8,46 GB
2. Klaster 1 : Segmen Entry-Level
Smartphone dalam klaster ini cenderung memiliki harga yang paling rendah, dengan RAM dan rating yang lebih rendah dibanding klaster lain. Cocok untuk pengguna dengan budget terbatas.
 - Rata-rata Harga: Rp2.565.787
 - Rata-rata Rating: 74,06
 - Rata-rata RAM: 5,16 GB
3. Klaster 2 : Segmen Mid-Range
Klaster ini menempati posisi tengah dari segi harga dan spesifikasi. Produk dalam klaster ini menawarkan nilai seimbang antara performa dan harga, cocok untuk pengguna yang menginginkan performa baik tanpa harga terlalu tinggi.
 - Rata-rata Harga: Rp5.114.491
 - Rata-rata Rating: 82,40
 - Rata-rata RAM: 7,64 GB

Evaluasi

Uji evaluasi clustering dilakukan menggunakan metode Silhouette Score, yang bertujuan untuk mengukur sejauh mana objek-objek dalam suatu cluster memiliki kemiripan satu sama lain (kohesi) dan seberapa jauh perbedaannya dengan objek-objek di cluster lain (separasi). Nilai Silhouette Score berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa hasil clustering sangat baik. Berdasarkan hasil perhitungan, diperoleh nilai Silhouette Score sebesar 0.6090. Nilai ini menunjukkan bahwa pemisahan antar cluster tergolong cukup baik, meskipun belum optimal. Artinya, sebagian besar data berada Prosiding dalam cluster yang tepat, namun masih terdapat beberapa data yang tumpang tindih antar cluster.

```
[ ] silhouette_scores = []
k_candidates = range(3, 7)
for k in k_candidates:
    kmeans = KMeans(n_clusters=k, random_state=42)
    labels = kmeans.fit_predict(X)
    score = silhouette_score(X, labels)
    silhouette_scores.append(score)
    print(f'k={k}, Silhouette Score={score:.4f}')

best_k_index = np.argmax(silhouette_scores)
best_k = k_candidates[best_k_index]
print(f'Jumlah cluster terbaik berdasarkan Silhouette Score adalah {best_k}')
```

↗ k=3, Silhouette Score=0.6090
 k=4, Silhouette Score=0.5498
 k=5, Silhouette Score=0.5461
 k=6, Silhouette Score=0.5346
 Jumlah cluster terbaik berdasarkan Silhouette Score adalah 3

Gambar 6: Silhouette Score

IV. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma K-Means Clustering efektif dalam mengelompokkan smartphone berdasarkan spesifikasi harga dan kapasitas RAM, sehingga menghasilkan segmentasi pasar yang terstruktur dalam tiga kelas utama: high-end, mid-range, dan entry-level. Pendekatan ini memberikan kontribusi penting dalam meningkatkan akurasi strategi pemasaran bagi produsen dan memudahkan konsumen dalam memilih produk sesuai kebutuhan mereka. Dengan demikian, penerapan teknik data mining ini dapat memperkuat pengelolaan pasar produk teknologi secara lebih sistematis dan berbasis data.

DAFTAR PUSTAKA

- [1] N. Afiasari, N. Suarna, and N. Rahaningsi, "Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means," *Jurnal SAINTEKOM*, vol. 13, no. 1, pp. 100–110, Mar. 2023, doi: 10.33020/saintekom.v13i1.402.
- [2] C. Hafidz Ardana *et al.*, "Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-means Clustering," 2024.
- [3] E. Febrianty, L. Awalina, and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail Optimizing Marketing Strategies with Customer Segmentation Using K-Means

- Clustering on Online Retail Transactions,” *Jurnal Teknologi dan Informasi (JATI)*, vol. 13, 2023, doi: 10.34010/jati.v13i2.
- [4] N. A. P. Haryandini, W. Prihartono, and . F., “ANALISIS SEGMENTASI PELANGGAN UNTUK MENYUSUN STRATEGI PROMOSI MENGGUNAKAN K-MEANS: STUDI KASUS DI PT XYZ,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5697.
 - [5] A. muliawan Nur, M. Saiful2, H. Bahtiar, and Muhammad Taufik Hidayat, “Penerapan Algoritma K-Means Clustering Dalam Mengelompokkan Smartphone Yang Rekomendasi Berdasarkan Spesifikasi,” *Infotek: Jurnal Informatika dan Teknologi*, vol. 7, no. 2, pp. 478–488, Jul. 2024, doi: 10.29408/jit.v7i2.26283.
 - [6] W. G. S. Parwita, “Pengujian Akurasi Sistem Rekomendasi Berbasis Content-Based Filtering,” *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, vol. 14, no. 1, p. 27, Feb. 2019, doi: 10.30872/jim.v14i1.1272.
 - [7] S. L. Achmad, A. Fauzi, R. Rahmat, and J. Indra, “SEGMENTASI PELANGGAN MENGGUNAKAN K-MEANS CLUSTERING DI TOKO RETAIL,” *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, vol. 7, no. 2, p. 736, Dec. 2024, doi: 10.37600/tekinkom.v7i2.1226.
 - [8] L. Rajput and S. Singh, *Customer Segmentation of E-commerce data using K-means Clustering Algorithm*. 2023. doi: 10.1109/Confluence56041.2023.10048834.
 - [9] A. Putri Riani, A. Voutama, and T. Ridwan, “Volume 6 ; Nomor 1,” *Januari*, pp. 164–172, 2023, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsk/index>
 - [10] F. Pramataning Dewi, P. Siwi Aryni, and Y. Umaidah, “Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran,” *MEI*, 2011.
 - [11] H. Haviluddin, S. J. Patandianan, G. M. Putra, N. Puspitasari, and H. S. Pakpahan, “Implementasi Metode K-Means Untuk Pengelompokkan Rekomendasi Tugas Akhir,” *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, vol. 16, no. 1, p. 13, Mar. 2021, doi: 10.30872/jim.v16i1.5182.
 - [12] S. Paembonan, H. Abduh, and K. Kunci, “Penerapan Metode Silhouette Coeficient Untuk Evaluasi Clustering Obat Clustering; K-means; Silhouette coeficient,” 2021. doi: 10.51557/pt_jiit.v6i2.659.
 - [13] Y. Yulianti, D. Y. Utami, N. Hikmah, and F. N. Hasan, “PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENGETAHUI MINAT CUSTOMER DI TOKO HIJAB,” *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 241–246, Sep. 2019, doi: 10.33480/pilar.v15i2.650.
 - [14] E. Novalia and A. Voutama, “Prediction of Rice Field Planted Area with CRISP-DM Using Classification and Regression Tree (Cart) Algorithms,” 2023.
 - [15] J. Hao and T. K. Ho, “Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language,” Jun. 01, 2019, *SAGE Publications Inc.* doi: 10.3102/1076998619832248.