

**IMPLEMENTATION OF K-MEANS ALGORITHM ALGORITHM FOR
CLUSTERING DATA MINING IN CLUSTERING DIVERSE LAPTOP BRANDS**

**IMPLEMENTASI ALGORITMA K-MEANS UNTUK CLUSTERING DATA
MINING DALAM PENGELOMPOKAN BERAGAM MEREK LAPTOP**

Fadhila¹, Elin Haerani²

Universitas Islam Negeri Sultan Syarif Kasim Riau^{1,2}

12050122312@students.uin-suska.ac.id¹, elin.haerani@uin-suska.ac.id²

ABSTRACT

Laptops have become electronic devices that are much needed and in demand, especially by students and IT departments in companies. Compared to desktop computers, laptops have a more compact, lightweight and power efficient design. One way to group laptop brands is to use the K-Means Clustering algorithm in data mining. This grouping aims to identify product potential and analyze data on laptops that consumers recommend buying. This research is useful for grouping laptop data and purchasing interest into structured clusters using the K-Means Clustering method. With this algorithm, data for various laptop brands was successfully resolved into three clusters: Cluster 0 (1027 data) with low Ghz and low laptop prices, Cluster 1 (216 data) with high Ghz and medium prices, and Cluster 2 (2604 data) with the conclusion that Ghz is medium and the price is high.

Keywords: Laptop, K-means, Clustering, Data Mining

ABSTRAK

Laptop telah menjadi perangkat elektronik yang sangat dibutuhkan dan diminati, terutama oleh mahasiswa dan departemen IT di perusahaan. Dibandingkan dengan komputer desktop, laptop memiliki desain yang lebih ringkas, ringan, dan hemat daya. Salah satu cara untuk mengelompokkan merek laptop adalah dengan menggunakan algoritma K-Means Clustering dalam data mining. Pengelompokan ini bertujuan untuk mengidentifikasi produk potensial dan menganalisis data laptop yang rekomendasi untuk dibeli konsumen. Penelitian ini bermanfaat untuk mengelompokkan data laptop dan minat pembelian ke dalam cluster terstruktur menggunakan metode K-Means Clustering. Dengan algoritma tersebut, data berbagai merek laptop berhasil dikelompokkan menjadi tiga cluster: Cluster 0 (1027 data) dengan Ghz kecil dan harga laptop yang murah, Cluster 1 (216 data) dengan Ghz tinggi dan harga yang sedang, dan Cluster 2 (2604 data) dengan kesimpulan Ghz sedang dan harga yang mahal.

Kata Kunci: Laptop, Kmeans, Clustering, Data Mining

PENDAHULUAN

Perkembangan teknologi informasi terus meningkat seiring berjalannya waktu, kebutuhan akan teknologi informasi sangat diutamakan saat ini khususnya penggunaan laptop. Laptop kini menjadi salah satu barang elektronik yang di minati dan sangat penting bagi masyarakat khususnya mahasiswa, terlebih bagian IT di perusahaan[1]. Laptop berbeda dari komputer desktop terutama dalam hal ukurannya yang lebih kecil, bobotnya yang lebih ringan, dan konsumsi dayanya yang lebih efisien. Namun, laptop tetap memiliki komponen yang serupa dengan komputer desktop, seperti prosesor, memori, dan media

penyimpanan[2]. Dengan perkembangan teknologi, berbagai merek laptop terus bermunculan, masing-masing menawarkan keunggulan dan spesifikasi yang beragam. Hal ini seringkali membuat konsumen kesulitan menentukan laptop yang paling sesuai dengan kebutuhan mereka, karena banyaknya pilihan merek dan model di pasaran. [3].

Salah satu cara dan metode yang tepat dapat diterapkan untuk mengelompokkan merk laptop yaitu menggunakan algoritma *K-Means Clustering* pada data mining. K-Means Clustering adalah metode yang digunakan untuk mengelompokkan sekumpulan objek berdasarkan atribut

atau karakteristik yang serupa dalam berbagai inventaris. K-Means, sebagai salah satu analisis kluster yang populer dalam data mining, merupakan teknik kuantisasi vektor yang sering diterapkan pada masalah pengelompokan, seperti menentukan kategori penjualan, mulai dari yang paling laku, kurang laku, hingga tidak laku[4].

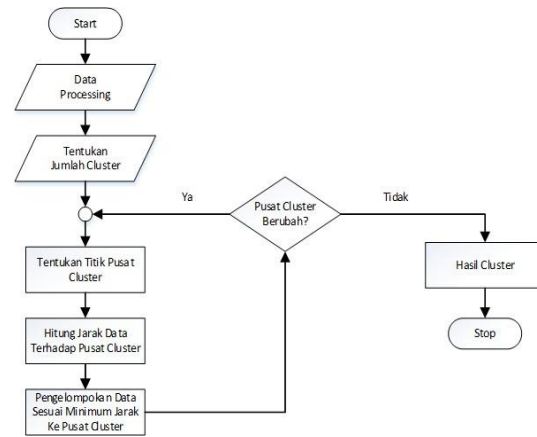
Penelitian sebelumnya yang dilakukan oleh Aji Supriyatno (2023) dengan judul “Analisis Segmentasi Pasar Laptop Dengan Metode Clustering dan Algoritma K-Means”[4] yang membahas tentang analisis segmentasi pasar yang membantu produsen dalam merancang produk yang sesuai dengan preferensi dan membantu konsumen dalam memberi kesimpulan dan manfaat potensial tentang laptop yang sesuai dengan kebutuhan.

Secara keseluruhan, tujuan dari pengelompokan data merk laptop adalah untuk menerapkan algoritma K-Means Clustering dalam mengidentifikasi produk yang berpotensi. Analisis data tingkat penjualan laptop dilakukan dengan menggunakan algoritma K-Means Clustering.[5]

Mengacu pada latar belakang diatas, maka penelitian ini menerapkan metode k-means clustering dalam pengelompokan data merk laptop untuk mengetahui hubungan yang positif dan signifikan antara persepsi konsumen terhadap merk, kualitas dengan minat pembeli

METODE PENELITIAN

Penelitian ini dilakukan dalam beberapa tahap yaitu:



Gambar 1. Flowchart Algoritma K-means

Gambar 1 diatas menunjukkan tahapan proses *clustering* dengan menggunakan algoritma *K-Means Clustering*. Jarak yang memisahkan dua objek digunakan untuk menentukan seberapa dekat kedua objek tersebut. Demikian pula, jarak antara suatu data dan pusat cluster digunakan untuk menentukan kedekatan data tersebut dengan cluster tersebut.

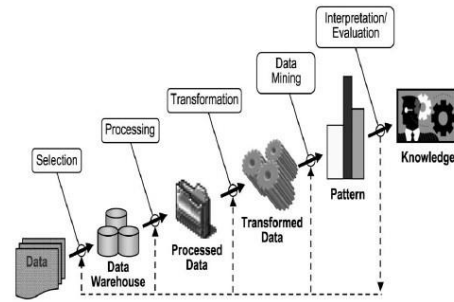
Pengumpulan Data

Pengumpulan data merupakan proses terstruktur untuk mengumpulkan informasi dari berbagai sumber yang sesuai dengan tujuan tertentu. Proses ini bertujuan untuk menjawab pertanyaan penelitian, mendukung pengambilan keputusan, atau memahami suatu fenomena secara lebih mendalam. Pengumpulan data dapat dilakukan melalui berbagai metode, seperti survei, wawancara, observasi, eksperimen, kajian dokumen, atau kombinasi dari beberapa metode tersebut. Pada penelitian ini jumlah data yaitu 3848 data dengan menggunakan 12 atribut dan 2 atribut khusus sebagai penentu yang dipakai dalam pengolahan data mining.

Tahapan KDD

Metode penelitian adalah proses menganalisis dan pemeriksaan data dan mengekstraksi data yang berguna. Metode yang diterapkan pada penelitian

ini yaitu knowledge Discovery in Database (KDD) yang terstruktur dalam 5 tahapan yaitu data selection (penyeleksi data), preprocessing atau cleaning (pembersih data), transformation (penyesuaian data), data mining (pencarian pola atau informasi), dan evaluation (penafsiran data). Tujuan utama dari Knowledge Discovery in Databases (KDD) adalah mengungkap pola, tren, serta pengetahuan tersembunyi dalam data yang dapat memberikan wawasan berharga untuk mendukung pengambilan keputusan. Proses KDD memanfaatkan beragam teknik dan algoritma data mining untuk mengekstrak informasi yang membantu meningkatkan pemahaman terhadap data, mengidentifikasi hubungan yang tidak terlihat secara langsung, dan menyusun strategi atau keputusan yang lebih efektif dalam berbagai bidang, seperti bisnis, sains, dan penelitian.[6]



Gambar 2. Tahapan KDD

Data Selection

Langkah awal dalam data mining adalah memilih data relevan dari kumpulan data operasional, dengan menyaring kolom atau baris berdasarkan kriteria tertentu untuk fokus pada informasi yang penting. Data yang ada pada data laptop ini tidak semua dipakai. Oleh karena itu hanya data yang sesuai untuk penelitian. Pada tahap seleksi data ini, atribut yang digunakan untuk proses data mining adalah Brand, Name, Price, Processor_Brand, RAM_Exp, RAM, Ghz, Display_Type, Display, GPU_Brand, SSD, Adapter. Adapun seleksi dalam penelitian ini dapat dilihat pada Tabel dibawah:

Tabel 1. Data Laptop

| BRAND | NAME | PRICE | PROCESSOR BRAND | RAM M | RAM EXP | GHZ | DISPLAY TYPE | DISPLAY Y | GPU BRAND | SSD | ADAPTER |
|--------|-----------------------------------|-------|-----------------|-------|---------|-----|--------------|-----------|-----------|-----|---------|
| LENOVO | LENOVO IDEAPAD SLIM3 (82KU017KN) | 36289 | AMD | 12 | 8 | 4.0 | LCD | 15.6 | AMD | 512 | 65 |
| DELL | DELLG15-5520 (D560822WIN9B) | 78500 | INTEL | 32 | 16 | 3.3 | LCD | 15.6 | NVIDIA | 512 | 56 |
| HP | HP 15S-FY5007TU (91R03PA) | 55490 | INTEL | 8 | 8 | 4.2 | LCD | 15.6 | INTEL | 512 | NO |
| ACER | ACER ONE 14 Z8-415 (UN.599SI.020) | 34990 | INTEL | 32 | 16 | 0 | LCD | 14 | INTEL | 512 | 45 |
| .. | .. | .. | | | ... | ... | ... | ... | ... | ... | ... |
| HP | HP14S-FQ1092AU (50M60PA) | 49599 | AMD | 16 | 8 | 4.0 | LCD | 14 | AMD | 512 | 65 |

Pada penelitian ini ada 12 atribut yang telah melalui tahapan preprocessing yang akan dipilih untuk melakukan pengelompokan clustering data mining. Dan memakai 2 atribut

husus yang paling relevan untuk pengelompokan hasil clustering k-means data laptop.

Tabel 2. Atribut

| No | Atribut |
|----|---------|
|----|---------|

| | |
|----|------------------|
| 1 | Brand |
| 2 | Name Data Laptop |
| 3 | Price |
| 4 | Processor_Brand |
| 5 | RAM_Exp |
| 6 | RAM |
| 7 | Ghz |
| 8 | Display_Type |
| 9 | Display |
| 10 | GPU_Brand |
| 11 | SSD |
| 12 | Adapter |

Tahapan pengumpulan data pada penelitian ini adalah pengumpulan data yang dimana para peneliti mencari data yang relevan dengan topik pembahasan dalam penelitian yang dilakukan. Proses pengumpulan data dilakukan dengan mencari data pada situs www.kaggle.com. Dengan keyword yaitu laptop dan dengan update terbaru sehingga dataset yang didapat sesuai dengan keadaan dan saat ini. Setelah dilakukan pencarian, peneliti menemukan sebuah dataset yang relevan dengan penelitian ini adalah data laptop dengan memakai 12 atribut yaitu Brand, Name, Price, Processor_Brand, RAM_Exp, RAM, Ghz, Display_Type, Display, GPU_Brand, SSD, Adapter. Pada penelitian ini terdapat jumlah data yaitu 3848 data. Dengan atribut khusus yang digunakan yaitu Price dan Ghz untuk mengelompokkan lebih spesifik terkait data laptop.

1. Data Preprocessing

Membersihkan data, merapikan dan mempersiapkan data agar dapat diolah dengan lebih efektif. Pada tahap ini Data preprocessing adalah langkah krusial dalam pipeline analitik atau pemodelan. Dengan data yang bersih, relevan, dan performa model machine learning dapat ditingkatkan secara signifikan.

2. Data Transformation

Sebuah proses yang dilakukan untuk

menyesuaikan data terpilih agar siap untuk data mining, meningkatkan kualitas data, mengatasi masalah tertentu, atau menjadikannya lebih relevan. Dalam memungkinkan data dapat diproses dengan K-Means clustering, data yang bersifat nominal perlu diinisialisasikan terlebih dahulu dalam bentuk angka. Karena dalam proses pengelompokan data yang dianalisis harus berupa data numerik. Adapun hasil dari tranformasi data pada penelitian ini dapat dilihat pada tabel dibawah ini:

Tabel 3. Hasil Inisialisasi Data

| Processor_Brand | Ket |
|-----------------|-----|
| AMD | 1 |
| Intel | 2 |
| Apple | 3 |
| Qualcomm | 4 |
| MediaTek | 5 |
| Microsoft | 6 |

3. Data Mining

Proses mengidentifikasi pola dan informasi menarik dalam data terpilih menggunakan metode tertentu, seperti k-means clustering, dengan mengelompokkan data ke dalam beberapa klaster berdasarkan kesamaan antar titik data. Pada proses K-means Clustering ini menggunakan 2 atribut penentu yaitu (Price, Ghz) dan proses Kmeans clustering nya adalah sebagai berikut

```
#menghitung silhouette score untuk nomer cluster
silhouette_scores = []
for n_clusters in range(3, 12):
    kmeans = KMeans(n_clusters=n_clusters, random_state=42)
    cluster_labels = kmeans.fit_predict(X_scaled)
    silhouette_avg = silhouette_score(X_scaled, cluster_labels)
    silhouette_scores.append(silhouette_avg)

# Plot skor silhouette scores
plt.figure(figsize=(10, 6))
plt.plot(range(2, 11), silhouette_scores, marker='o', linestyle='--')
plt.title('Silhouette Scores untuk jumlah cluster yang berbeda')
plt.xlabel('Jumlah cluster')
plt.ylabel('Silhouette Score')
plt.xticks(range(2, 11))
plt.grid(True)
plt.show()
```

Gambar 3. Codingan Clustering Data Mining

Selanjutnya dengan menggunakan matrik Silhouette Score untuk mengevaluasi hasil clustering data mining

```

from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import silhouette_score
import warnings

warnings.filterwarnings("ignore")

#pilih colum untuk clustering
X = df[['Ghz', 'Price']]

# standardize the data
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
    
```

Gambar 4. Silhouette_Score

Interpretasi/Evaluasi

Evaluasi algoritma mencakup memahami, Pola informasi yang dihasilkan dari data mining harus disajikan dalam format yang mudah dipahami oleh pihak terkait. Interpretasi difokuskan pada menghubungkan kluster ke konteks nyata (kategori pasar

laptop), sementara evaluasi menggunakan metrik seperti Silhouette Score untuk memastikan kluster yang dihasilkan berkualitas dan bermakna.

HASIL DAN PEMBAHASAN PENELITIAN

Dalam pengujian analisis metode K-Means Clustering dalam sistem pengelompokan data, diperlukan data sebagai input untuk proses dan analisis. Setelah melakukan preprocessing data, data yang mendukung penelitian ini untuk dianalisis menggunakan metode K-Means Clustering adalah sebagai berikut:

Tabel 4. Data Laptop

| No | Brand | Name | Price | Process orBrand | Ram Exp | RAM | GH Z | Displ ay_T ype | .. | .. | SSD | Adapter |
|------|--------|--|-------|-----------------|---------|-----|------|----------------|-----|-----|-----|---------|
| 1 | Lenovo | Lenovo Ideapad Slim 3 (82KU017KN) AMD Hexa Core Ryzen 5 | 36289 | AMD | 12 | 8 | 4.0 | LCD | .. | .. | 512 | 65 |
| 2 | Dell | Dell G15-5520 (D560822WIN9B) Core i5 12th Gen | 78500 | Intel | 32 | 16 | 3.3 | LCD | .. | .. | 512 | 56 |
| 3 | HP | HP 15s-fy5007TU (91R03P) Core i5 12th Gen | 55490 | Intel | 8 | 8 | 4.2 | LCD | .. | .. | 512 | No |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 3848 | Lenovo | Lenovo Legion 5i (82AU | 69990 | Intel | Not | 8 | 2.5 | LED | .. | .. | 512 | 60 |

00CFI
N)

Dalam tahap ini langkah-langkah clustering pengelompokan data mining adalah sebagai berikut:

1. Pemodelan

Pada tahap pembahasan yang dilakukan, akan dijelaskan tahap dalam merancang pemodelan. Pemodelan ini dimulai dari tahapan proses pemilihan library yang akan digunakan dari proses data preprocessing, data mining serta pengujian data.

2. Library Pemodelan

Library atau modul merupakan kumpulan fungsi, alat, dan kelas yang dibuat untuk mendukung pengembang dalam merancang model matematika, statistik, atau simulasi untuk berbagai aplikasi. Library ini umumnya dimanfaatkan dalam pengembangan perangkat lunak yang membutuhkan analisis data, machine learning, atau representasi sistem kompleks dari dunia nyata.

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
from sklearn.metrics import silhouette_score
sns.set_theme(color_codes=True)
pd.set_option('display.max_columns', None)
```

Gambar 5. Library

3. K-Means Clustering

K-Means Clustering adalah teknik clustering yang sederhana namun kuat untuk banyak aplikasi analisis data. Algoritma ini sangat efektif untuk dataset dengan klaster yang terdefinisi dengan baik, meskipun memerlukan perhatian ekstra terhadap parameter K, pemilihan centroid awal, dan skala data. Dengan interpretasi yang tepat, K-Means dapat memberikan wawasan strategis dari dataset kompleks dan membantu dalam pengambilan keputusan berbasis data. Berikut

adalah hasil cluster dari clustering data mining:

Hasil Clustering Data Laptop
Tabel 5. Hasil Cluster

| No | Merek Laptop | Cluster |
|----|--|---------|
| 1 | Lenovo Ideapad Slim 3 (82KU017KIN) Laptop (15.6 Inch AMD Hexa Core Ryzen 5 8 Windows 11 512 SSD) | C2 |
| 2 | Dell G15-5520 (D560822WIN9B) Laptop (15.6 Inch Core i5 12th Gen 16 Windows 11 512 SSD) | C0 |
| 3 | HP 15s-fy5007TU (91R03PA) Laptop (15.6 Inch Core i5 12th Gen 8 Windows 11 512 SSD) | C2 |
| 4 | Acer One 14 Z8-415 (UN.599SI.020) Laptop (14 Inch Core i5 11th Gen 16 Windows 11 512 SSD) | C2 |
| 12 | Lenovo LOQ 15APH8 (82XT004JIN) Laptop (15.6 Inch AMD Octa Core Ryzen 7 16 Windows 11 512 SSD) | C1 |
| 15 | Acer Predator Helios Neo 16 PHN16-71 (NH.QLTSI.003) Laptop (16 Inch Core i7 13th Gen 16 Windows 11 1 TB SSD) | C1 |
| 16 | HP 14s-fq1092AU (50M60PA) Laptop (14 Inch | C0 |

4. Berikut merupakan visualisasi data Scatterplot

Visualisasi data Scatterplot merupakan jenis visualisasi data yang menunjukkan hubungan antara dua variabel. Setiap titik pada scatterplot merepresentasikan sebuah observasi atau data, dengan letaknya ditentukan oleh nilai kedua variabel tersebut pada sumbu horizontal (x) dan vertikal (y).



Gambar 6. Visualisasi Data

Secara umum, Scatterplot berguna untuk mengeksplorasi hubungan,

pola, atau anomali pada dataset sebelum analisis lebih lanjut, titik yang jauh dari kelompok utama mengindikasikan data yang tidak biasa. Setelah mendapatkan Label cluster untuk masing-masing data $n=1,2,3,4,5,\dots$ maka dicari nilai rata-rata dengan menjumlahkan masing-masing cluster dan dibagi jumlah anggotanya. Dan dari tabel dibawah didapatkan hasil klaster dengan memakai atribut khusus sebagai berikut:

Tabel 6. Hasil Cluster

| Cluster | GHZ | PRICE |
|---------|-----|---------------------------|
| C0 | 2.4 | 5.175.400- 7.472.420 |
| C1 | 3.3 | 23.378,100- 18.518,500 |
| C2 | 3.2 | 28.067,700- 36.607,100 |

Nilai rata-rata centroid pada pengulangan ke-2, karena centroid tidak mengalami perubahan (sama dengan centroid sebelumnya) maka proses clustering selesai. Jadi dari tabel diatas didapatkan penjelasan sebagai berikut: Hasil C0 menghasilkan 1027 data dengan kesimpulan bahwa C0 adalah laptop dengan Ghz kecil dan harga yang murah

Hasil C1 menghasilkan 216 data dengan kesimpulan bahwa C1 adalah laptop dengan Ghz tinggi dengan harga yang sedang

Hasil C2 menghasilkan 2604 data dengan kesimpulan bahwa C2 adalah laptop dengan Ghz sedang dengan harga yang mahal

PENUTUP

Kesimpulan

Pada Penelitian ini diperoleh kesimpulan yaitu tidak semua atribut dalam data memiliki kontribusi yang signifikan untuk menentukan kelompok dan 2 atribut yang dipilih mungkin

adalah atribut yang paling relevan atau berpengaruh berdasarkan analisis sebelumnya. Untuk atribut Harga dan Ghz mungkin lebih relevan untuk pengelompokan karena konsumen cenderung membandingkan perangkat berdasarkan ini. melalui algoritma K-Means Clustering pada pengelompokan data beragam merek laptop berhasil dikelompokkan menjadi tiga cluster. Cluster (C0) terdiri dari 1027 data dengan Ghz kecil dan harga yang murah, cluster 1 (C1) terdiri dari 216 data dengan Ghz tinggi dan harga yang sedang dan cluster 2 (C2) terdiri dari 2604 data dengan Ghz sedang dan harga yang relatif mahal. Hasil dari penelitian ini, dapat digunakan sebagai panduan bagi konsumen untuk memilih laptop yang sesuai dengan preferensi dan kebutuhan.

Saran dari penelitian ini, penulis merekomendasikan untuk melanjutkan penelitian ini dengan menerapkan algoritma lain dari data mining, seperti yang penulis sarankan memakai algoritma DBSCAN karna algoritma ini kana algoritma ini sangat efektif mendeteksi cluster dengan bentuk arbitrer (tidak harus bulat). Algoritma ini juga fleksibilitas dan ketahanannya terhadap nilai noise menjadikannya pilihan utama untuk berbagai aplikasi di data mining.

DAFTAR PUSTAKA

- [1] U. N. Putra *et al.*, "Penerapan data mining K-Means clustering untuk mengelompokkan berbagai jenis merk leptop," *Semin. Nas. Sist. ...*, pp. 241–249, 2021, [Online]. Available: <https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/35%0Ahttps://sismatik.nusaputra.ac.id/index.php/sismatik/article/download/35/31>
- [2] A. Putri Riyandoro, A. Voutama, and

- Y. Umaidah, “Implementasi Data Mining Clustering K-Means Dalam Menggolongkan Beragam Merek Laptop,” *JATI (Jurnal Mhs. Tek. Inform.,* vol. 7, no. 2, pp. 1372–1377, 2023, doi: 10.36040/jati.v7i2.6816.
- [3] W. I. Rahayu, S. F. Pane, L. A. Frimanda, and K. Kunci, “IMPLEMENTASI DATA MINING DENGAN METODE K-MEANS CLUSTERING UNTUK MENENTUKAN IKLAN AUDIO BERDASARKAN USER BEHAVIORS PADA APLIKASI AUDIO SOCIAL MEDIA SVARA DI PT . ZAMRUD KHATULISTIWA TECHNOLOGY,” vol. 10, no. 2, pp. 13–19, 2018.
- [4] I. R. A. K. J. R. P. D. M. Aji Supriyatno, “Analisis Segmentasi Pasar Laptop Dengan Metode Clustering Dan Algoritma K-Means,” no. June, pp. 2–22, 2023, [Online]. Available: <https://www.researchgate.net/publication/371675029>
- [5] N. Firdaus, S. Defit, and G. Widi Nurcahyo, “Penerapan Algoritma K-Means Clustering Dalam Pengelompokan Data Penjualan Pada Pabrik Mobil Toyota Indonesia,” *Fountain Informatics J.,* vol. 8, no. 2, pp. 2548–5113, 2023, [Online]. Available: <http://dx.doi.org/10.21111/fij.v8i2.10798>
- [6] V. Alvianatinova, I. Ali, N. Rahaningsih, and A. Bahtiar, “Penerapan Algoritma K- Means Clustering Dalam Pengelompokan Data Penjualan Supermarket Berdasarkan Cabang (Branch),” *JATI (Jurnal Mhs. Tek. Inform.,* vol. 8, no. 2, pp. 1529–1535, 2024, doi: 10.36040/jati.v8i2.8993.