Implementation Of Data Mining In Determining Car Sales Strategy Using The K-Means Algorithm And Clustering K-Medoids [Implementasi Data Mining Dalam Menentukan Strategi Penjualan Mobil Menggunakan Algoritma K-Means Dan K-Medoids Clustering]

Harisma Firdaus Hibatullah[1)], Uce Indahyanti[2)], Yulian Findawati[3)], Suprianto[4)]

[1,2,3,4)] Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia
Email Penulis Korespondensi: uceindahyanti@umsida.ac.id

Abstract. *The demand for transportation modes in today's society continues to increase, leading many people to choose to own private vehicles as a means to support their daily activities. Cars have become one of the most preferred modes of transportation among the Indonesian people. This research aims to analyze car sales data by applying the K-Means and K-Medoids algorithms to identify public preferences for transmission types and to formulate sales strategies. The clustering results indicate that the K-Means algorithm is optimal at K=2 with a DBI value of 0.492, dividing the data into low and high sales groups. Meanwhile, K-Medoids shows the best performance at K=5 with a DBI of 0.573, resulting in more detailed segmentation based on transmission dominance and sales volume. These findings indicate that K-Medoids is superior in generating specific strategic information. The results of this study are expected to serve as a reference in formulating car sales strategies that are more adaptive to market changes.*

Keywords – *Data Mining; K-Means; K-Medoids; Sales*

Abstrak *Kebutuhan masyarakat pada moda transportasi masa kini terus mengalami peningkatan, sehingga banyak masyarakat memilih untuk memiliki kendaraan pribadi sebagai sarana penunjang aktivitas sehari-hari. Mobil menjadi salah satu pilihan transportasi yang paling diminati oleh masyarakat Indonesia. Penelitian ini bertujuan untuk menganalisis data penjualan mobil dengan menerapkan algoritma K-Means dan K-Medoids guna mengidentifikasi preferensi masyarakat terhadap jenis transmisi, serta merumuskan strategi penjualan. Hasil clustering menunjukkan bahwa algoritma K-Means optimal pada K=2 dengan nilai DBI sebesar 0,492, membagi data ke dalam kelompok penjualan rendah dan tinggi. Sementara itu, K-Medoids menunjukkan performa terbaik pada K=5 dengan DBI sebesar 0,573, menghasilkan segmentasi lebih rinci berdasarkan dominasi transmisi dan volume penjualan. Temuan ini menunjukkan bahwa K-Medoids lebih unggul dalam menghasilkan informasi strategis yang spesifik. Hasil penelitian ini diharapkan dapat menjadi referensi dalam penyusunan strategi penjualan mobil yang lebih adaptif terhadap perubahan pasar.*

Kata Kunci – *Data Mining; K-Means; K-Medoids; Penjualan*

I. PENDAHULUAN

Kebutuhan masyarakat pada moda transportasi masa kini terus mengalami peningkatan, sehingga banyak masyarakat memilih untuk memiliki kendaraan pribadi sebagai sarana penunjang aktivitas sehari-hari. Mobil menjadi salah satu pilihan transportasi yang paling diminati oleh masyarakat Indonesia. [1]. Penjualan mobil di Indonesia sempat mengalami pertumbuhan pada tahun 2022 dengan total penjualan wholesales sebesar 1.049.040 unit dan penjualan retail sebesar 1.013.582 unit. Tetapi, pada tahun 2023 penjualan mobil di Indonesia terjadi penurunan dengan total penjualan wholesales sejumlah 1.005.802 unit dan retail sejumlah 998.059 unit. [2].

Pada tahun 2024 penjualan mobil kembali mengalami penurunan dengan target awal 1 juta unit menjadi 850 ribu dan berhasil terjual sebanyak 865 ribu. Perusahaan otomotif terus berupaya untuk mencari tahu merk kendaraan apa yang paling diminati oleh masyarakat Indonesia, dengan tujuan dapat memenuhi kebutuhan konsumen sekaligus memprediksi penjualan di masa depan. Satu dari sekian banyak pendekatan yang dapat diterapkan dalam mengatasi permasalahan tersebut ialah dengan menerapkan data mining melalui teknik clustering.

Data mining atau Knowlegde Discovery in Databases (KDD) ialah suatu aktivitas yang berhubungan dengan mengolah data guna memperoleh pola atau ikatan yang nantinya dapat digunakan dalam pengambilan keputusan [3]. Clustering merupakan teknik yang umum digunakan data mining dengan membagi sekumpulan data menjadi cluster

atau kelompok tertentu. Data dengan ciri yang serupa dikelompokkan menjadi satu cluster dan data dengan ciri berbeda akan dialokasikan pada cluster lain [4].

Beberapa penelitian yang berhubungan dengan implementasi K-Means dan K-Medoids yang pernah diterapkan sebelumnya, yaitu: Pengelompokan potensi produksi kelapa sawit di Indonesia yang mngacu pada luas area, produksi, dan produktivitas. Penelitian tersebut menghasilkan tiga cluster: rendah (18 wilayah), sedang (5 wilayah), dan tinggi (2 wilayah), dengan potensi tertinggi terdapat di Kalimantan Barat dan Riau. K-Means digunakan dalam perhitungan rata-rata luas area (514.885,72 Ha), produksi (1.931.882,84 ton), dan produktivitas (3.227,08 kg/Ha), sementara K-Medoids memperkuat segmentasi. Hasilnya menunjukkan 72% wilayah berpotensi rendah, 20% sedang, dan 2% tinggi, dengan kombinasi algoritma yang terbukti efektif dan akurat, sesuai dengan perhitungan manual [5].

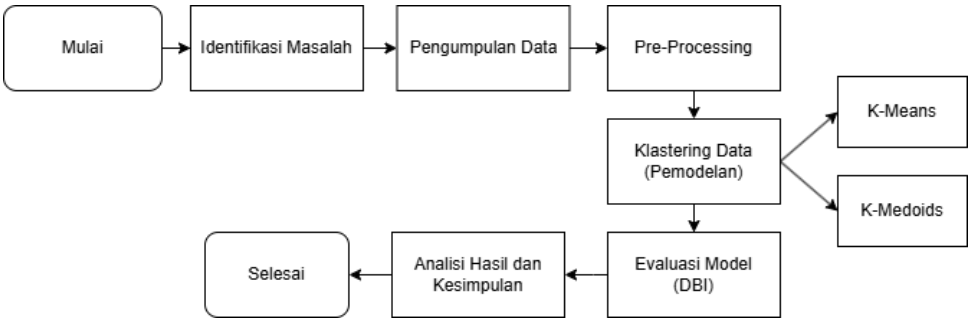
Pada penelitian lain, perbandingan antara K-Means dan K-Medoids untuk mengelompokan wilayah produksi kakao di Sulawesi Selatan yang menampilkan nilai Davies-Bouldin pada K-Means berada dibawah K-Medoids sebesar 0,292 pada nilai k = 4. Hasil pengelompokan tersebut menghasilkan cluster C0 dengan 17 daerah, C1 dengan 5 daerah, serta C2 dan C3 masing-masing berisi 1 daerah. Hasil tersebut menyatakan bahwa K-Means lebih optimal dari K-Medoids [6].

Pada penelitian lainnya, yang dilakukan dengan menerapkan K-Means dalam mengelompokkan data penjualan mobil dalam kurun waktu 5 tahun, dengan total 900 data. Terdapat 3 kelompok utama, yaitu: Cluster 0 yang diberi label "Laris" dengan 235 anggota (26%), Cluster 1 yang dikategorikan "Kurang Laris" dengan 604 anggota (67%), dan Cluster 2 yang disebut "Paling Laris" dengan 61 anggota (7%). Dengan nilai DBI sebesar 0,341. [1].

Berdasarkan latar belakang diatas, maka diusulkan penelitian ini dengan menggunakan dataset 5 tahun terakhir, dengan menggunakan 4 atribut utama yaitu bulan penjualan mobil, merek mobil, jenis transmisi dan jumlah mobil yang laku terjual, serta menambahkan metode k-medoids yang dapat digunakan sebagai metode pembanding dari penelitian sebelumnya. Juga, diharapkan temuan hasil penelitian ini dapat berfungsi sebagai panduan untuk penelitian pada tahun berikutnya.

II. METODE

Metode penelitian mencakup serangkaian langkah sistematis, dimulai dengan identifikasi masalah hingga analisis hasil dan kesimpulan yang dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

Identifikasi Masalah

Perbedaan kinerja antara algoritma K-Means dan K-Medoids telah dibahas dalam berbagai studi, dengan hasil yang bervariasi. Penelitian ini bertujuan untuk menerapkan kedua algoritma *clustering* pada data penjualan mobil guna mengetahui merek dengan jenis transmisi yang paling diminati masyarakat. Selain itu, penelitian ini juga bertujuan mengevaluasi hasil *clustering* untuk memberikan rekomendasi strategi penjualan yang lebih efektif.

Pengumpulan Data

Langkah selanjutnya, dilakukan tinjauan pustaka dengan mengumpulkan berbagai rujukan dari jurnal yang relevan dengan topik penelitian [7]. Selanjutnya, mengumpulkan beberapa sampel data yang tersedia pada situs resmi Gabungan Industri Kendaraan Bermotor Indonesia (https://www.gaikindo.or.id/).

Pre-Processing

Data yang telah diperoleh akan dilakukan seleksi atribut, pengelompokkan data, dan normalisasi data. Tujuan dilakukannya *pre-processing* ialah untuk memudahkan berjalannya proses *clustering* data dan menghindari dominasi dari fitur tertentu yang memiliki skala nilai yang lebih besar [8].

Klastering Data

Data mining ialah tahapan untuk melakukan identifikasi terhadap pola yang unik dari sekumpulan data yang dipilih dengan menerapkan algoritma yang sejalan dengan tujuan analisis [9]. Penelitian ini melakukan perbandingan performa dari K-Means dan K-Medoids pada pengolahan data penjualan mobil sebagai langkah awal dalam menentukan strategi penjualan.

1) K-Means

K-Means clustering ialah metode mesin tanpa pengawasan (*unsupervised machine learning*) yang berfungsi sebagai pembagi dataset menjadi k kelompok atau *cluster*. Teknik ini mengelompokkan data berdasarkan kesamaan karakteristik yang diwakili oleh centroid atau titik pusat [10]. Tahapan melakukan *clustering* dengan metode K-Means adalah sebagai berikut :

1. Tentukan jumlah *cluster k*
2. Pilih titik pusat (centroid) awal secara acak.
3. Hitung jarak setiap titik terhadap centroid dengan rumus jarak Euclidean yang dirumuskan pada Persamaan 1.

$$d = \sqrt{\sum_N (x_i + y_i)^2} \quad (1)$$

Dimana:

d = Jarak antar data.

x_i = Atribut pada data pertama.

y_i = Atribut pada data kedua.

4. Mengelompokkan setiap titik dengan centroid terdekat.
5. Perbarui centroid dengan menghitung rata-rata dari semua titik pada data untuk setiap cluster.
6. Ulangi langkah 3 sampai 5, hingga nilai centroid konvergen (tidak berubah secara signifikan).

2) K-Medoids

K-Medoids atau Partitioning Around Medoids (PAM) ialah salah satu bentuk model K-Means yang memilih pusat klaster (medoid) dengan cara acak dari titik data dalam *cluster*. Pendekatan ini bertujuan untuk meminimalkan pengaruh outlier dan mengurangi sensitivitas terhadap klaster yang dihasilkan [11]. Tahapan melakukan clustering dengan metode K-Medoids adalah sebagai berikut :

1. Tentukan jumlah *cluster k.*
2. Pilih titik pusat (medoid) awal secara acak.
3. Hitung jarak setiap titik terhadap medoid dengan menggunakan rumus jarak Euclidean yang dirumuskan pada Persamaan 1.
4. Kelompokkan setiap titik dengan medoid terdekat
5. Hitung nilai cost (S), yang dirumuskan pada Persamaan 2.

$$S = B - A \quad (2)$$

Dimana:

A = Total jarak terdekat titik dengan medoid awal

B = Total jarak terdekat titik dengan medoid sementara.

6. Ulangi langkah 3 sampai 5 hingga nilai S > 0

Evaluasi Model

David L. Davies dan Donald W. Bouldin berhasil menciptakan sebuah model evaluasi di tahun 1979, yang diberi nama Davies Bouldin Index (DBI) guna menilai kualitas dari perhitungan algoritma clustering. DBI menggunakan skema evaluasi internal untuk menilai kualitas klaster berdasarkan kuantitas dan kedekatan antar data dalam klaster. [12]. Klaster yang optimal ditandai dengan nilai kohesi yang tinggi, yaitu data pada cluster mempunyai persamaan karakteristik yang tinggi dan separasi yang rendah, yang menunjukkan perbedaan yang jelas antar klaster. Nilai DBI yang lebih rendah atau mendekati nol, menunjukkan kualitas *cluster* yang lebih baik [4]. Berikut merupakan tahapan untuk menghitung evaluasi DBI:

1. Hitung nilai kohesi dengan menggunakan rumus *Sum of Square Within* (SSW) yang dirumuskan pada Persamaan 3.

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (3)$$

Dimana:

m_i = Jumlah data pada c*luster* ke-i.

$d(x_j, c_i)$ = Jarak data x_j dengan *cluster ke*-i.

2. Hitung nilai separasi dengan menggunakan rumus *Sum of Square Between* yang dirumuskan pada Persamaan 4.

$$SSB_{i,j} = d\ (c_i, c_j) \quad (4)$$

Dimana:
c_i, c_j = Titik pusat cluster ke-i dan ke-j.

3. Hitung nilai rasio untuk mengetahui perbandingan dari nilai kohesi dan separasi yang dirumuskan pada Persamaan 5.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (5)$$

4. Evaluasi kualitas clustering dilakukan dengan Davies-Bouldin Index (DBI) yang dirumuskan pada Persamaan 6.

Cluster yang baik ditandai dengan nilai DBI yang paling rendah hingga mendekati 0.

$$DBI = \frac{1}{k}\sum_{i=0}^{k} R_i \quad (6)$$

Dimana:
k = Jumlah *cluster*.

Analisis Hasil dan Kesimpulan

Pada tahapan ini dilakukan analisis terhadap pola yang dihasilkan dari perhitungan algoritma k-means dan k-medoids hingga evaluasi disajikan dalam bentuk visualisasi guna memberikan gambaran yang lebih jelas serta mendukung kesimpulan akhir dari penelitian.

III. HASIL DAN PEMBAHASAN

Pengumpulan Data dan Pre-Processing

Dataset yang digunakan berjumlah 1.452 data hasil penjualan mobil dari tahun 2020 hingga 2024. Pengambilan data dilakukan berdasarkan rekap penjualan yang diperoleh dari website resmi GAIKINDO, dan dikonversi ke dalam format file CSV. Atribut yang digunakan dalam penelitian ini terdiri dari Bulan, Brand, Transmisi dan Total [13]. Dataset penjualan mobil ditampilkan pada Tabel 1.

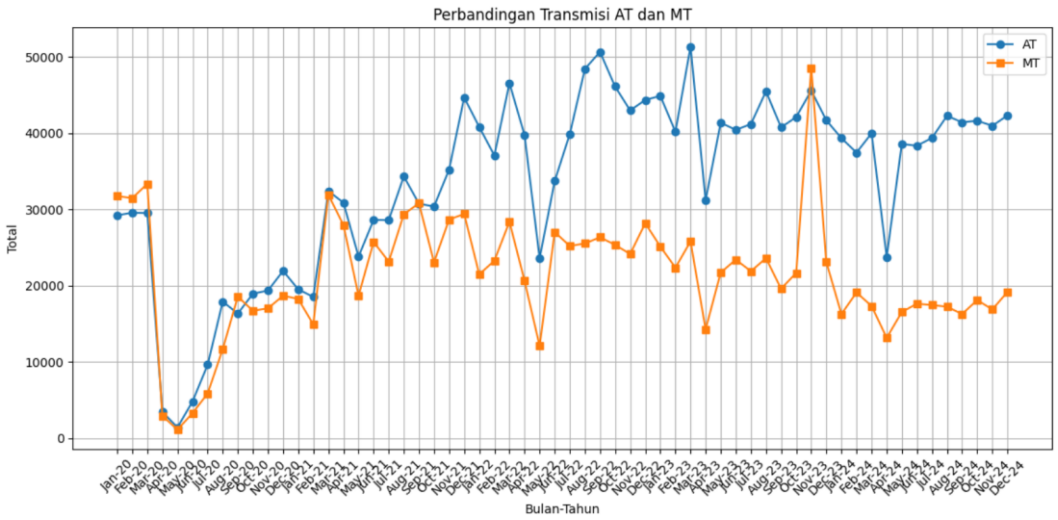
Table 1. Dataset

Bulan	Brand	Transmisi	Total
Jan - 20	HONDA	AT	8812
Jan – 20	HONDA	MT	3248
Jan – 20	HYUNDAI	AT	130
…	…	…	…
Dec – 24	VOLVO CARS	MT	0
Dec – 24	TANK	AT	89
Dec – 24	TANK	MT	0

Langkah selanjutnya, dataset dilakukan pengelompokkan berdasarkan Brand, Bulan_Tahun, AT dan MT. Pengelompokkan dilakukan agar format data menjadi lebih terstruktur sehingga identifikasi lebih akurat. Hasil pengelompokkan data ditampilkan pada Tabel 2 dan Gambar 2.

Table 2. Pengelompokkan Data

Brand	Bulan_Tahun	AT	MT
HONDA	Jan – 20	8812	3248
HYUNDAI	Jan – 20	130	19
MERCEDES BENZ	Jan – 20	123	0
…	…	…	…
ISUZU	Dec – 24	0	0
VOLVO CARS	Dec – 24	0	0
TANK	Dec – 24	8	0



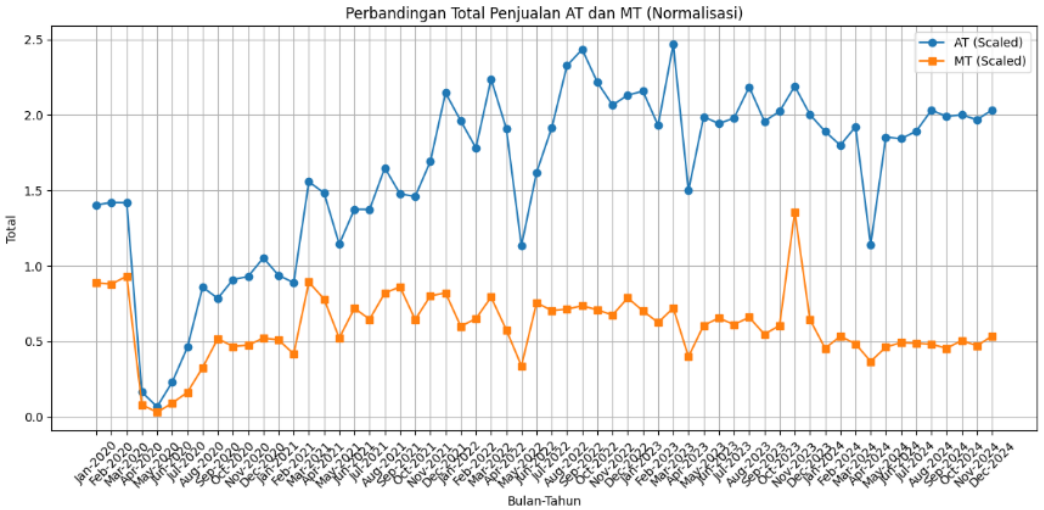
Gambar 2. Perbandingan Transmisi

Pada Gambar 2, grafik perbandingan penjualan memperlihatkan adanya fluktuasi. Oleh karena itu, perlu dilakukan normalisasi data menggunakan Min-Max Normalization guna menyetarakan skala penjualan menjadi skala 0 hingga 1. Hasil dari normalisasi data ditampilkan pada Tabel 3 dan Gambar 3.

Tabel 3. Normalisasi Data

Brand	Bulan_Tahun	AT	MT
HONDA	Jan – 20	0.423	0.091
HYUNDAI	Jan – 20	0.006	0.001
MERCEDES BENZ	Jan – 20	0.006	0.000
…	…	…	…
ISUZU	Dec – 24	0.000	0.000

VOLVO CARS	Dec – 24	0.000	0.000
TANK	Dec – 24	0.004	0.000



Gambar 3. Perbandingan Transmisi (Normalisasi)

Klastering Data

Pada tahapan ini, pengujian dilakukan dengan jumlah cluster yang bervariasi mulai 2 hingga 10 cluster. Percobaan dilakukan untuk menemukan jumlah cluster yang paling optimal pada algoritma K-Means maupun K-Medoids.

K-Means

Tahapan awal proses klasterisasi data menggunakan algoritma K-Means, ialah melakukan perhitungan untuk mengetahui jarak data dengan titik pusat awal (*centroid*) yang dipilih secara acak dengan menggunakan rumus jarak *Euclidean*. Berikut tahapan untuk perhitungan pada *cluster* 2 dengan algoritma K-Means:

5. Menentukan titik pusat (centroid) awal secara acak. *Centroid* awal ditampilkan Tabel 4.

Tabel 4. Centroid Awal

Cluster	AT_scaled	MT_scaled
0	0.005	0.000
1	0.643	0.249

2. Hitung jarak setiap titik terhadap centroid dengan Persamaan 1.

$$d = \sqrt{\sum_N (x_i + y_i)^2} \quad (1)$$

Hitung jarak data ke 1 dengan centroid ke 0

$\sqrt{(0.423 - 0.005)^2 + (0.091 - 0.000)^2}$

= 0.428

Hitung jarak data ke 1 dengan centroid ke 1

$\sqrt{(0.002 - 0.643)^2 + (0.000 - 0.249)^2}$

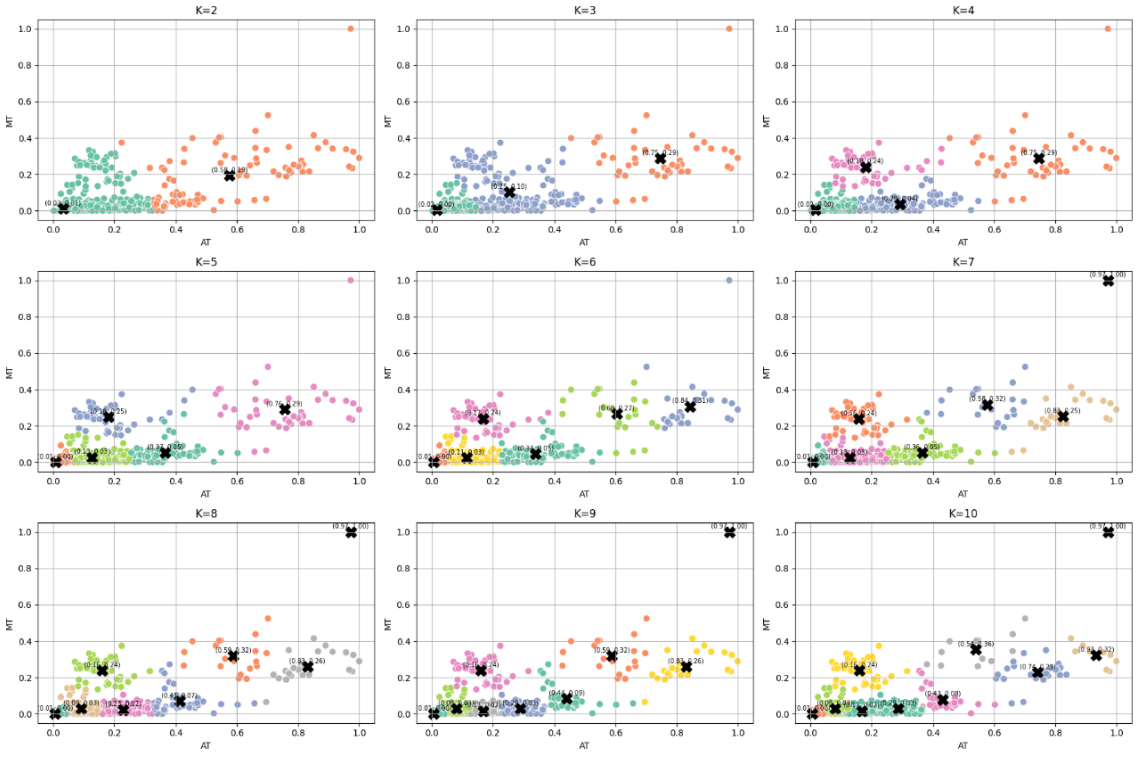
= 0.271

3. Kelompokkan setiap titik data berdasarkan cluster terdekat, dapat dilihat pada Tabel 5.

Tabel 5. Hasil Perhitungan Iterasi 1

AT_scaled	MT_scaled	Jarak Terdekat	Cluster
0.428	0.271	0.271	1
0.001	0.683	0.001	0
0.001	0.684	0.001	0
…	…	…	…
0.005	0.689	0.005	0
0.005	0.689	0.005	0
0.001	0.686	0.001	0

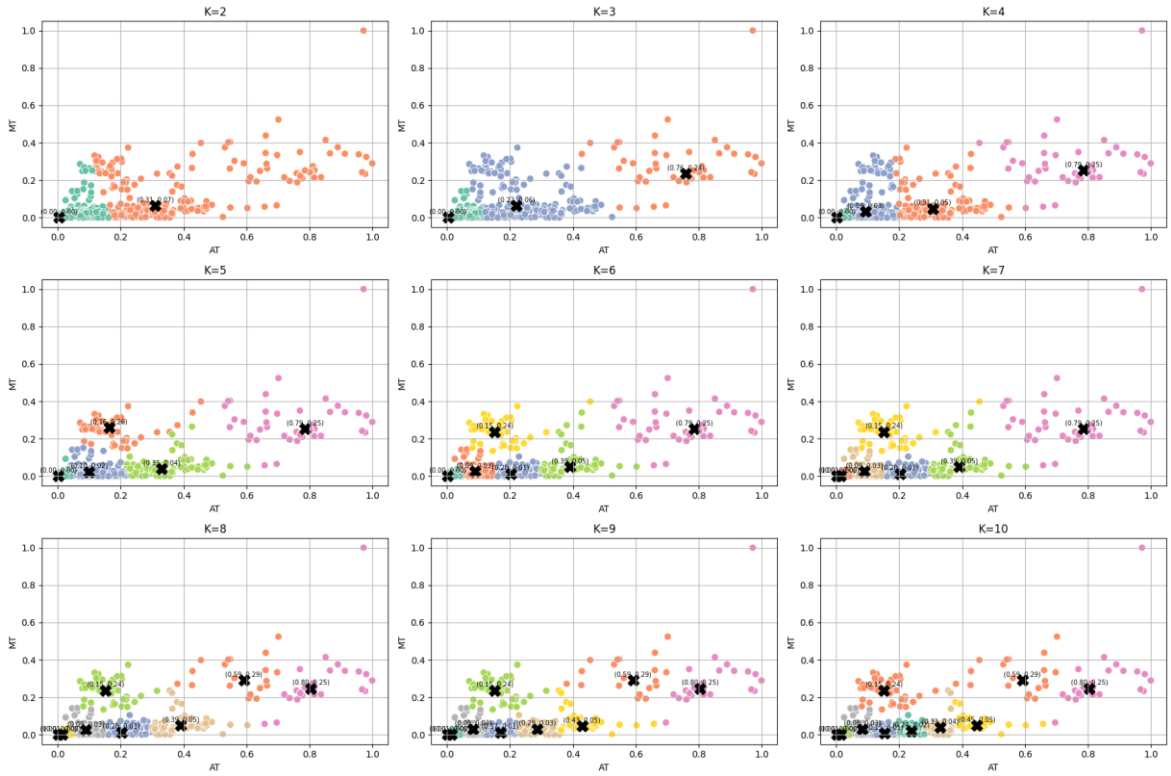
4. Perbarui centroid dengan menghitung rata-rata dari semua titik data pada setiap cluster.
5. Ulangi kembali langkah 3 hingga centroid yang baru konvergen (tidak berubah secara signifikan)
6. Lakukan perhitungan serupa untuk cluster 3 hingga 10.
   Hasil dari seluruh percobaan dengan menggunakan pendekatan K-Means ditampilkan pada Gambar 4.



Gambar 4. Klasterisasi K-Means

K-Medoids

Pada algoritma K-Medoids dilakukan percobaan yang serupa. Namun, berbeda dengan K-Means yang menghentikan iterasi saat *centroid* konvergen, K-Medoids menghentikan iterasi saat nilai cost (S > 0) atau ketika pemilihan medoid baru tidak lagi menghasilkan pengurangan nilai cost. Hasil klasterisasi dari algoritma K-Medoids ditampilkan pada Gambar 5.



Gambar 5. Klasterisasi K-Medoids

Evaluasi Model

Tahap selanjutnya ialah melakuan pengujian kualitas terhadap pembentukan *cluster* untuk menentukan cluster yang paling optimal dengan menggunakan model evaluasi DBI. Tahapan dari perhitungan evaluasi DBI untuk cluster 2 pada algoritma K-Means sebagai berikut :

Tabel 6. Centroid Akhir K-Means

Cluster	AT_scaled	MT_scaled
0	0.032	0.013
1	0.576	0.192

1. Hitung nilai kedekatan dalam 1 *cluster* dengan menggunakan rumus Persamaan 3.

   $SSW_i = \frac{1}{m_i}\sum_{j=1}^{m_i} d(x_j, c_i)$ (3)

   Kedekatan titik data pada *cluster* 0

   $SSW_0 = \frac{0.2598913+0.25016185+\cdots+0.04421388}{1358} = 0.050374666$

   Kedekatan titik data pada *cluster* 1

   $SSW_1 = \frac{0.39644+0.23945+\cdots+0.20879}{94} = 0.231300092$

2. Hitung jarak antar cluster dengan menggunakan rumus Persamaan 4.

   $SSB_{i,j} = d(c_i, c_j)$ (4)

   Jarak antara *cluster* 0 dengan *cluster* 1

   $SSB_{0,1} = \sqrt{(0.032-0.576)^2 + (0.13-0.192)^2} = 0.572829965$

3. Hitung rasio dengan menggunakan rumus Persamaan 5.

   $R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$ (5)

Rasio antara kohesi dan separasi

$$R_{0,1} = \frac{0.050374666+0.231300092}{0.572829965} = 0.491724902$$

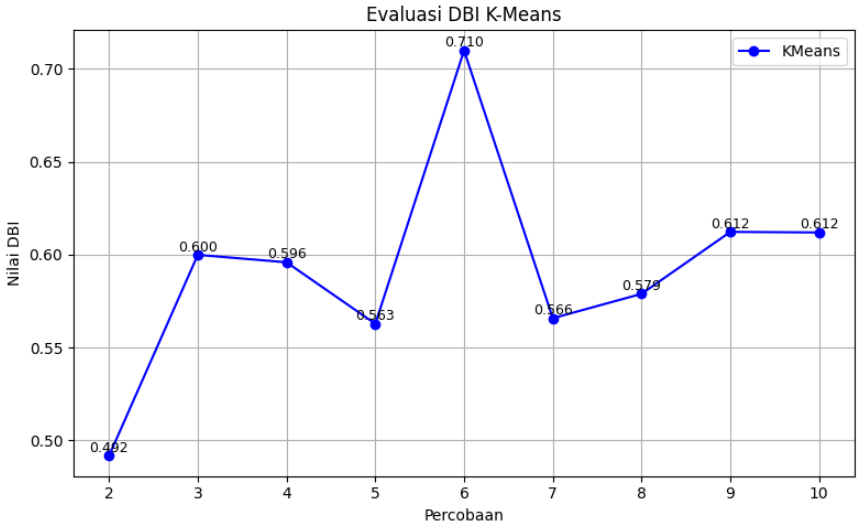
4. Hitung nilai DBI menggunakan rumus Persamaan 6.

$$DBI = \frac{1}{k}\sum_{i=0}^{k} R_i \quad (6)$$

Dikarenakan cluster hanya berjumlah 2, maka :

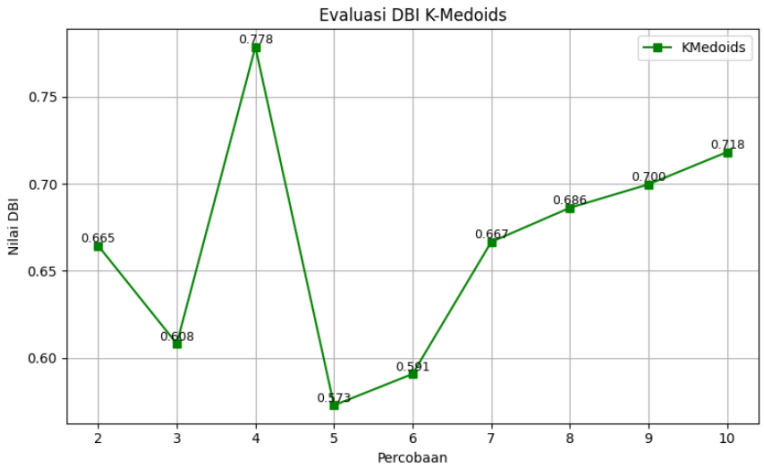
$$DBI = \frac{R_0+R_1}{2} = \frac{2.R_{0,1}}{2} = R_{0,1} = 0.491724902$$

5. Ulangi langkah 1 hingga 4 untuk cluster 3 hingga 10.
Hasil dari seluruh pengujian evaluasi pada algoritma K-Means ditampilkan pada Gambar 6.



Gambar 6. Evaluasi K-Means

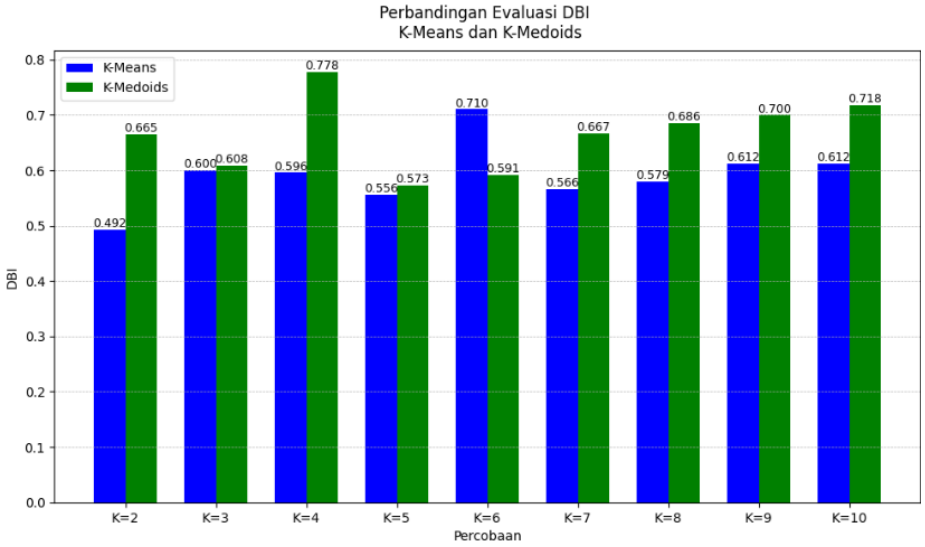
Dilanjutkan dengan menguji performa algoritma K-Medoids dengan menggunakan metrik yang sama. Evaluasi ini dilakukan untuk membandingkan efektivitas kedua algoritma dalam membentuk klaster yang optimal berdasarkan karakteristik data penjualan mobil. Hasil dari seluruh pengujian evaluasi untuk algoritma K-Medoids ditampilkan pada Gambar 7.



Gambar 7. Evaluasi K-Medoids

Analisis Hasil

Setelah dilakukan serangkaian percobaan dengan jumlah *cluster* yang bervariasi, mulai dari 2 hingga 10 cluster, langkah selanjutnya adalah melakukan analisa terhadap hasil dari pembentukan *cluster* dari algoritma K-Means dan K-Medoids, serta nilai evaluasi terbaik dari masing-masing algoritma sebagai bagian dari kesimpulan dari penelitian. Hasil perbandingan dari evaluasi DBI dari kedua algoritma ditampilkan pada Gambar 8.



Gambar 8. Perbandingan Evaluasi DBI

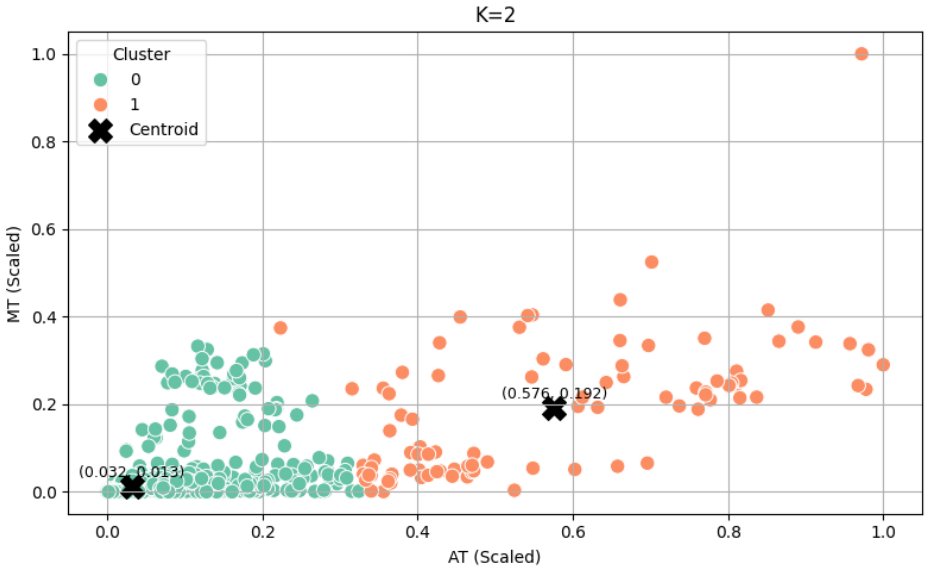
Pada Gambar 11 menunjukkan K-Means mencapai performa optimal pada jumlah cluster K = 2 dengan nilai evaluasi DBI sebesar 0,492. Dengan distribusi data sebanyak 1.358 data pada Cluster 0, yang mewakili kelompok dengan penjualan rendah (AT dan MT) dan Cluster 1 dengan 94 data, mewakili kelompok dengan tingkat penjualan tinggi (AT dan MT). Strategi yang dapat diterapkan untuk Cluster 0 dapat difokuskan pada peningkatan pemasaran dan diskon untuk memperluas pangsa pasar. Cluster 1 dapat difokuskan untuk mempertahankan performa penjualan melalui peningkatan layanan purna jual dan ketersediaan stok. Hasil klasterisasi K-Means ini disajikan pada Tabel 7, Tabel 8, dan Gambar 9.

Tabel 7. *Cluster* 0 K-Means.

Rata-Rata AT	Rata-Rata MT	Keterangan	Brand
662.2	465.1	Rendah (AT dan MT)	AUDI, BMW, DFSK, HYUNDAI, ISUZU, KIA, LEXUS, MAZDA, MERCEDES BENZ, MINI, NISSAN, PEUGEOT, SUZUKI, VAKSWAGON, WULING, DAIHATSU, MITSUBISHI MOTORS, MORRIS GARAGE, SUBARU, HONDA, CHERRY, NETA, RENAULT, SERES, AION, BAIC, BYD, CHEVROLET, CITROEN, DATSUN, FORD, HAVAL JEEP, VOLVO CARS, TANK, TOYOTA

Tabel 8. *Cluster* 1 K-Means

Rata-Rata AT	Rata-Rata MT	Keterangan	Brand
11989.4	6882.01	Tinggi (AT dan MT)	TOYOTA, HONDA, MITSUBISHI MOTORS, DAIHATSU



Gambar 9. K-Means

Sementara itu, K-Medoids menunjukkan performa optimal pada jumlah *cluster* K = 5 dengan nilai evaluasi DBI sebesar 0,573. Dengan distribusi data pada Cluster 0 terdiri dari 51 data yang mewakili kelompok dengan penjualan rendah-menengah (Dominasi MT), Cluster 1 terdiri dari 81 data yang mewakili kelompok penjualan tinggi (AT dan MT), Cluster 2 terdiri dari 46 data yang mewakili kelompok penjualan sangat tinggi (AT dan MT), Cluster 3 terdiri dari 1.095 data yang mewakili kelompok penjualan rendah (AT dan MT), dan Cluster 4 terdiri 179 data yang mewakili kelompok penjualan rendah-menengah (Dominasi AT).

Strategi yang dapat diterapkan pada hasil klasterisasi algoritma k-medoids yaitu, Cluster 0 diperlukan stretegi pemasaran yang efektif dan melakukan edukasi *value* transmisi AT, Cluster 1 dan 2 difokuskan untuk memaksimalkan distribusi dan pemasaran terhadap keunggulan produk, serta memberikan layanan purna jual, Cluster 3 diperlukan evaluasi terhadap produk yang kurang diminati dan *niche* yang lebih sempit, Cluster 4 diperlukan stretegi pemasaran yang efektif dan melakukan edukasi *value* transmisi MT. Hasil klasterisasi K-Medoids ditampilkan pada Tabel 9, Tabel 10, Tabel 11, Tabel 12, Tabel 13, dan Gambar 10.

Tabel 9. *Cluster* 0 K-Medoids

Rata-Rata AT	Rata-Rata MT	Keterangan	Brand
3756.3	8920	Rendah-Menengah (Dominasi MT)	DAIHATSU, TOYOTA

Tabel 10. *Cluster* 1 K-Medoids

Rata-Rata AT	Rata-Rata MT	Keterangan	Brand
7080.5	1755.4	Tinggi (AT dan MT)	HONDA, HYUNDAI, MITSUBISHI MOTORS, NISSAN, TOYOTA

Tabel 11. *Cluster* 2 K-Medoids

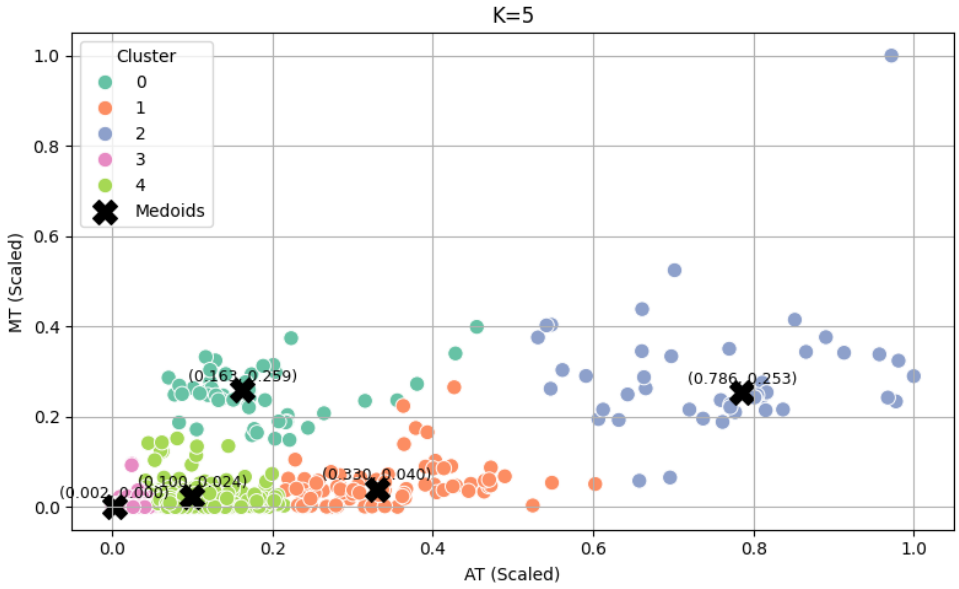
Rata-Rata AT	Rata-Rata MT	Keterangan	Brand
15736.4	10424.3	Sangat Tinggi (AT dan MT)	HONDA, TOYOTA

Tabel 12. *Cluster* 3 K-Medoids

Rata-Rata AT	Rata-Rata MT	Keterangan	Brand
123.3	28.6	Rendah (AT dan MT)	AUDI, BMW, PEUGEOT, MINI, VOLKSWAGON, MERCEDES BENZ, MAZDA, LEXUS, KIA, ISUZU, DFSK, NISSAN, MORRIS GARAGE, SUBARU, WULING, SERES, RENAULT, NETA, HYUNDAI, CHERY, VOLVO CARS, TANK, JEEP, FORD, HAVAL, DATSUN, CITROEN, CHEVROLET, BAIC, AION, SUZUKI, DAIHATSU, BYD, MITSUBISHI MOTORS, HONDA, TOYOTA

Tabel 13. *Cluster* 4 K-Medoids

Rata-Rata AT	Rata-Rata MT	Keterangan	Brand
2247	952.6	Rendah-Menengah (Dominasi AT)	SUZUKI, MITSUBISHI MOTORS, HYUNDAI, WULING, DAIHATSU, BYD, HONDA, TOYOTA, CHERRY, NISSAN



Gambar 10. K-Medoids

IV. SIMPULAN

Penelitian ini berhasil menerapkan algoritma *clustering* K-Means dan K-Medoids untuk menganalisis data penjualan mobil, dengan tujuan utama mengidentifikasi preferensi masyarakat terhadap jenis transmisi serta merumuskan strategi penjualan Hasil pengolahan data menggunakan dua algoritma *clustering* menunjukkan bahwa algoritma K-Means mencapai performa optimal pada jumlah klaster K = 2, dengan nilai Davies-Bouldin Index (DBI) sebesar 0,492. Clustering ini menghasilkan segmentasi data dalam skala makro, yaitu membedakan antara kelompok dengan volume penjualan rendah hingga menengah dan kelompok dengan penjualan tinggi. Sementara itu, algoritma K-Medoids menunjukkan hasil terbaik pada jumlah klaster K = 5 dengan nilai DBI sebesar 0,573, yang mampu menghasilkan segmentasi dalam skala mikro secara lebih rinci. Segmentasi tersebut mencakup berbagai karakteristik kelompok merek berdasarkan dominasi jenis transmisi dan tingkat volume penjualan yang lebih bervariasi. Perbandingan kedua hasil tersebut menunjukkan bahwa K-Means lebih efektif digunakan untuk analisis dalam skala luas, sedangkan K-Medoids lebih unggul dalam mendukung pengambilan keputusan strategis yang lebih terfokus, seperti inovasi produk, penguatan citra merek, hingga strategi pemasaran yang lebih terarah.

REFERENSI

[1] S. Butsianto and N. T. Mayangwulan, "Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering," J. Nas. Komputasi dan Teknol. Inf., vol. 3, no. 3, pp. 187–201, 2020, doi: 10.32672/jnkti.v3i3.2428.

[2] D. F. RIYANDA, "POLA PEMBELIAN MOBIL SUZUKI BERDASARKAN LOKASI DAN JENISNYA MENGGUNAKAN ALGORITMA K-MEANS." UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU, 2022.

[3] N. T. Luchia, H. Handayani, F. S. Hamdi, D. Erlangga, and S. F. Octavia, "Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia: Comparison of K-Means and K-Medoids on Poor Data Clustering in Indonesia," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 2, no. 2, pp. 35–41, 2022.

[4] E. T. E. Tasia, "Perbandingan Algoritma K-Means Dan K-Medoids Untuk Clustering Daerah Rawan Banjir Di Kabupaten Rokan Hilir: Comparison Of K-Means And K-Medoid Algorithms For Clustering Of Flood-Prone Areas In Rokan Hilir District," Indones. J. Inform. Res. Softw. Eng., vol. 3, no. 1, pp. 65–73, 2023.

[5] F. H. Pratama, A. Triayudi, and E. Mardiani, "Data mining k-medoids dan k-means untuk pengelompokan potensi produksi kelapa sawit di indonesia," JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform., vol. 7, no. 4, pp. 1294–1310, 2022.

[6] N. A. S. Z. Abidin, R. D. Avila, A. Hermatyar, and R. Rismayani, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Produksi Kakao," J. Tek. Inform. dan Sist. Inf., vol. 8, no. 2, pp. 383–391, 2022.

[7] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," J. Ilm. Teknol. dan Rekayasa, vol. 25, no. 1, pp. 76–88, 2020.

[8] R. Gustrianda and D. I. Mulyana, "Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma K-Means Dan K-Medoids," J. Media Inform. Budidarma, vol. 6, no. 1, p. 27, 2022.

[9] M. Kadafi, "Penerapan algoritma fp-growth untuk menemukan pola peminjaman buku perpustakaan UIN raden fatah Palembang. Matics, 10 (2), 52." 2019.

[10] D. Aggarwal and D. Sharma, "Application of clustering for student result analysis," Int J Recent Technol Eng, vol. 7, no. 6, pp. 50–53, 2019.

[11] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. I. R. H. Zer, and D. Hartama, "Analisis algoritma k-medoids clustering dalam pengelompokan penyebaran covid-19 di indonesia," J. Teknol. Inf., vol. 4, no. 1, pp. 166–173, 2020.

[12] N. K. Zuhal, "Study Comparison K-Means Clustering Dengan Algoritma Hierarchical Clustering: AHC, K-Means Clustering, Study Comparison," in Seminar Nasional Teknologi & Sains, 2022, pp. 200–205.

[13] E. Rahmah, "Penerapan Algoritma K-Medoids Clustering untuk menentukan Strategi Promosi Pada Data Mahasiswa (Studi Kasus: STIKES PERINTIS PADANG)," Penerapan Algoritm. K-Medoids Clust. untuk menentukan Strateg. Promosi Pada Data Mhs. (Studi Kasus STIKES PERINTIS PADANG), vol. 5, no. 03, pp. 556–564, 2022.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.