

IMPLEMENTASI ALGORITMA CLUSTERING PADA DATA PENJUALAN MOBIL DI AUTO2000 SALEMBA

Fericho Rusman, Vitri Tundjungsari
Teknik Informatika, Universitas Esa Unggul
Jl Arjuna Utara No.9 Jakarta, Indonesia
Fericho31@gmail.com

ABSTRAK

Dalam industri otomotif, analisis data menjadi alat penting untuk memahami pola penjualan dan preferensi pelanggan guna mengembangkan strategi pemasaran yang efektif. Auto2000 Salemba, salah satu cabang PT Astra International Tbk, belum menerapkan metode analisis berbasis data untuk mengelompokkan penjualan mobil berdasarkan atribut seperti varian, warna, jenis transmisi, dan harga. Penelitian ini bertujuan untuk mengelompokkan mobil terlaris berdasarkan variannya di Auto2000 Salemba, Jakarta, Indonesia. memberikan wawasan tentang segmentasi pasar dan posisi kompetitif. Metode K-means diterapkan menggunakan data penjualan yang dikumpulkan dari Auto2000 Salemba, cabang dari PT. Astra International Tbk. Dataset ini terdiri dari 1501 catatan penjualan mobil, di mana atribut seperti varian mobil, warna, jenis transmisi, dan harga dianalisis. Prosesnya melibatkan pembersihan data, transformasi, dan penerapan algoritma K-Means untuk mengidentifikasi kluster yang mengkategorikan varian mobil terlaris. Melalui analisis, jumlah kluster optimal (k) ditentukan sebanyak 3, yang dipilih berdasarkan Metode Elbow. Hasil menunjukkan bahwa Kluster 1 terdiri dari 474 kendaraan dengan preferensi warna tertentu, Kluster 2 mencakup 889 kendaraan dengan volume penjualan tinggi, dan Kluster 3 berisi 138 kendaraan dengan preferensi yang lebih beragam. Dari hasil yang didapatkan pada penelitian ini dapat membantu Auto2000 mengembangkan strategi pemasaran.

Kata kunci: *Clustering, K-Means, Auto2000 Salemba.*

1. PENDAHULUAN

Mobil merupakan salah satu alat transportasi utama di perkotaan karena memberikan kenyamanan, kecepatan, dan kemudahan dalam perjalanan. Dengan tingginya persaingan di industri otomotif, penting bagi dealer mobil seperti Auto2000 Salemba untuk memahami pola penjualan guna mengembangkan strategi pemasaran yang efektif. Berdasarkan wawancara dengan pihak manajemen Auto2000 Salemba, diketahui bahwa analisis berbasis data sangat diperlukan untuk mendukung strategi pemasaran dan memahami preferensi pelanggan secara lebih mendalam dan Auto2000 Salemba belum menerapkan metode analisis data untuk mengelompokkan penjualan mobil berdasarkan atribut spesifik seperti varian, warna, dan harga.

Penelitian ini bertujuan untuk membantu pihak Auto2000 dengan mengelompokkan varian mobil berdasarkan data penjualan menggunakan algoritma K-Means. Algoritma K-Means telah digunakan secara luas dalam berbagai studi untuk segmentasi data, termasuk pada konteks otomotif. Salah satunya adalah penelitian terdahulu yang menggunakan metode K-Means untuk menentukan strategi penjualan mobil berdasarkan atribut spesifik dalam Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil [1].

Penelitian ini diharapkan memberikan wawasan terkait preferensi pelanggan untuk mendukung pengambilan keputusan strategis, termasuk pengembangan strategi pemasaran yang lebih efektif.

2. TINJAUAN PUSTAKA

2.1. Data mining

Tahap Data Mining mengikuti analisis sumber spesifik untuk mendapatkan informasi berharga. Informasi berguna ini diperoleh melalui beberapa proses kompleks seperti kecerdasan buatan atau kecerdasan buatan, matematika, pembelajaran mesin, dll. Metode sulit berikut selanjutnya mengekstrak informasi berguna dari data besar. Data mining adalah teknik analisis data yang dapat mengekstrak informasi penting dari kumpulan data yang sangat besar. Data mining ialah proses inti pada Knowledge Discovery in Databases (KDD) karena proses ini menentukan arah atau bentuk pengetahuan akhir. KDD merupakan fungsi yang digunakan untuk mencari informasi atau data yang terdapat pada suatu dataset [2].

2.2. Clustering

Pengelompokan data berdasarkan kesamaan di antara data dikenal sebagai clustering. Metode ini mengelompokkan data dengan atribut yang serupa ke dalam cluster yang sama, sementara data dengan atribut yang berbeda dikelompokkan ke dalam cluster yang berbeda [3].

2.3. K-Means

Dengan mengadopsi K-means Clustering, dimungkinkan guna mencari data yang serupa atau serupa dan mengorganisasikan data ke dalam kelompok-kelompok data lainnya. K-means adalah kombinasi atau pengelompokan data di seluruh kelompok perhentian yang telah ditentukan. Tujuan dari metode K adalah untuk mengumpulkan data dari

beberapa kelompok yang memiliki karakteristik bermacam-macam dan menjadi kelompok yang berbeda dengan pengetahuan yang ada [4]. Dengan menggunakan K-means, data dapat diurutkan ke dalam satu atau lebih tanpa pengelompokan hierarki. Dengan menggunakan pendekatan pembelajaran langsung, algoritma K-Means dapat mengatasi masalah meminimalkan kesalahan duplikat. Untuk mengimplementasikan algoritma K-Means, pertama-tama harus dibuat K cluster, menugaskannya secara acak ke nilai centroid tertentu, menentukan jarak data setiap sentropot, dan menggunakan metode korelasi jarak Euclidean antara dua objek [5], yang dimana rumus nya sebagai berikut:

$$Shi = \sqrt{(m_h - m_i)^2 + (n_h - n_i)^2 + \dots (q_h - q_i)^2} \quad (1)$$

Dengan penjelasan sebagai berikut:

S_{hi} = Jarak antara h dan i

m_h = Koordinat m objek

m_i = Koordinat m pusat

n_h = Koordinat n objek

Setelah menghitung jarak antar masing-masing data, langkah selanjutnya adalah menghubungkan seluruh data observasi berdasarkan jarak terdekat antara cluster dengan data tersebut[6].

Langkah selanjutnya adalah mencari centroid baru dengan merata-ratakan data dari cluster yang dibandingkan [7].

Untuk mencari centroid baru, gunakan rumus berikut untuk menghitung jarak antara data dan centroid:

$$Kc = \frac{s1+s2+s3+\dots+sn}{\sum s} \quad (2)$$

Dengan penjelasan sebagai berikut:

Kc = data cluster ke c

$s1$ = nilai record data ke 1

$s2$ = nilai record data ke 2

$\sum s$ = jumlah record data.

2.4. Machine Learning

Kecerdasan buatan memiliki bidang tersendiri yang disebut *Machine Learning* (ML) yang berkaitan dengan perolehan informasi dari data, yaitu mengembangkan sistem yang dapat belajar secara "mandiri" tanpa perlu diprogram manusia lagi [8].

2.5. Metode Waterfall

Waterfall adalah model dokumentasi yang disukai selama proses pembuatan proyek di perkantoran atau industri besar. Model ini dapat dimanfaatkan untuk proyek yang membutuhkan material berkualitas tinggi [9].

2.6. PHP Hypertext Preprocessors

PHP merupakan kependekan dari *HyperText Preprocessor*, bahasa skrip sisi server yang tertanam

di HTML dan dijalankan di server. Selain berfungsi sebagai alat eksekusi skrip server, PHP juga dapat dimanfaatkan untuk membuat aplikasi desktop. Dampak dari hal ini adalah PHP memiliki potensi untuk menjadi target serangan dari pihak yang tidak diinginkan [10].

3. METODE PENELITIAN

Pada penelitian ini, analisis menggunakan Algoritma K-Means dengan memanfaatkan data sekunder dari Auto2000 Salemba. Proses analisis ini mencakup tahap pengumpulan data, analisis data, dan penerapan Algoritma K-Means untuk mengidentifikasi pola yang ada. Penelitian ini akan dilaksanakan dalam beberapa tahap, yang meliputi langkah-langkah sebagai berikut:

3.1. Analisa Masalah

Pada tahap ini, peneliti melakukan analisis terhadap masalah yang didapatkan dan melakukan studi literatur pada penelitian yang relevan.

3.2. Pengumpulan Data

Pada tahap pengumpulan data, peneliti mendapatkan data sekunder dari Auto2000 sebanyak 1501 data penjualan mobil, adapun atribut yang dianalisis meliputi varian mobil, warna (Color Desc), jenis transmisi dan harga.

3.3. Pra-Pemrosesan Data

Pada tahap ini dilakukan preprocessing untuk meningkatkan kualitas dan keakuratan data. Tahap preprocessing termasuk pembersihan, penggantian nilai yang hilang, normalisasi, dan transformasi untuk memastikan bahwa data yang akan digunakan konsisten.

3.4. Penerapan Algoritma K-Means

Pada tahap ini diterapkannya Algoritma k-means yang merupakan algoritma pengelompokan yang membagi data ke dalam beberapa kelompok berdasarkan kemiripan karakteristik. Pada tahap ini, tujuan utama adalah guna mencari data yang serupa dan mengorganisasikan data ke dalam kelompok-kelompok data lainnya. Proses penggunaan algoritma K-Means pertama-tama harus dibuat K cluster, menugaskannya secara acak atau mengambil nilai rata-rata ke nilai centroid tertentu, menentukan jarak data setiap centroid, dan menggunakan metode korelasi jarak Euclidean antara dua objek. Setelah menghitung jarak antar masing-masing data, langkah selanjutnya adalah menghubungkan seluruh data observasi berdasarkan jarak terdekat antara cluster dengan data tersebut. Langkah selanjutnya adalah mencari centroid baru dengan merata-ratakan data dari cluster yang dibandingkan.

3.5. Implementasi

Pada tahap ini, sistem dirancang menggunakan metode waterfall dan bahasa pemrograman PHP untuk

mengimplementasikan algoritma K-Means. Desain mencakup antarmuka pengguna yang dimulai dengan halaman login admin, fitur impor dataset dalam format Excel, dan pengaturan kluster.

4. HASIL DAN PEMBAHASAN

4.1. Data Hasil Penelitian

Pada penelitian ini menggunakan data sekunder dari Auto2000 Salemba sebanyak 1501, berikut merupakan data yang telah dilakukan pembersihan data dan transformasi:

Tabel 1. Dataset

No	Typo Kendaraan	Color Desc.	Total AR
1.	W01QVA/T02	Platinum White Pearl Mica	275.600.000
2.	A250SA/T20	BLACK,	280.800.000
3.	GUN166VRZA/T21	SUPER WHITE II	572.700.000
4.	W01QVA/T00	SILVER METALLIC..	277.600.000
5.	W01QVA/T02	Platinum White Pearl Mica	275.600.000
6.	A250SA/T21	Black Silver Metalic	258.400.000
7.	W01GM/T00	SILVER METALLIC..	230.100.000
8.	A250SA/T20	DARK GREY MICA	280.800.000
9.	W01QVA/T02Y	Platinum White Pearl Mica	285.392.000
10.	MAG10QHA/T00	ATTITUDE BLACK MICA	611.000.000

4.2. Pembahasan Hasil Penelitian

Penelitian ini menggunakan algoritma K-Means untuk mengelompokkan data penjualan mobil Auto2000 Salemba berdasarkan Color Desc, Transmisi Mesin, dan Harga. Berdasarkan analisis data penjualan menggunakan metode elbow, dapat

digunakan tiga kluster utama. Perhitungan jarak antar data dilakukan dengan Euclidean Distance, yang mencapai kestabilan pada iterasi ketujuh sebagai berikut:

Tabel 2. Hasil Perhitungan

No	Nama Kendaraan (Model)	C1	C2	C3	Hasil Cluster
1	W01QVA/T02 (VELOZ 1.5 Q CVT premium color), Platinum White Pearl Mica	13.0	2.16	29.41	cluster2
2	A250SA/T20Y (RAIZE LUX 1.0T GR SPORT CVT ONE TONE LUX), BLACK	2.52	9.58	18.42	cluster1
3	GUN166VRZA/T21 (FORTUNER 2.8 VRZ GR-S 4x2 A/T), SUPER WHITE II	2.47	12.58	15.50	cluster1
4	W01QVA/T00 (VELOZ 1.5 Q CVT (NonPremium Color), SILVER METALLIC	1.6	11.56	16.43	cluster1
5	W01QVA/T02 (VELOZ 1.5 Q CVT Premium Color), Platinum White Pearl Mica	13.00	2.16	29.41	cluster2
6	A250SA/T21 (RAIZE LUX 1.0T GR SPORT CVT TWO TONE LUX), Black Silver Metalic	10.03	2.10	26.41	cluster2
7	W01GM/T00 (AVANZA 1.5 G M/T), SILVER METALLIC	3.43	12.07	16.82	cluster1
8	A250SA/T20Y (RAIZE LUX 1.0T GR SPORT CVT ONE TONE LUX), DARK GREY MICA	1.89	10.57	17.43	cluster1
9	W01QVA/T02Y (VELOZ 1.5 Q CVT LUX Premium Color), Platinum White Pearl Mica	13.00	2.16	29.41	cluster2
10	MAG10QHA/T00 (INNOVA ZENIX HYBRID 2.0 Q HV MODELISTA CVT TSS NPC), ATTITUDE BLACK MICA	12.09	1.81	28.45	cluster2

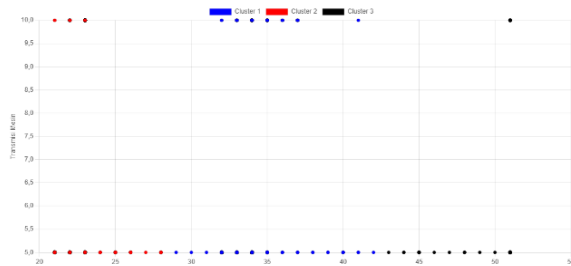
Dalam proses pengelompokan data menggunakan algoritma K-Means, dilakukan perhitungan secara manual dan sistematis untuk memastikan hasil yang dihasilkan. Hasil pengelompokan menunjukkan bahwa terdapat tiga cluster utama yang dapat diidentifikasi berdasarkan karakteristik data yang dianalisis. Cluster 1 terdiri dari 474 kendaraan, Cluster 2 mencakup 889 kendaraan, dan Cluster 3 mencakup 138 kendaraan. Masing-masing cluster ini mencerminkan preferensi pelanggan yang berbeda, khususnya dalam hal warna, jenis

transmisi mesin, dan rentang harga kendaraan. Informasi ini memberikan wawasan yang dapat dimanfaatkan oleh Auto2000 Salemba untuk lebih memahami kebutuhan dan preferensi pelanggan. Gambaran detail hasil perhitungan ini disajikan pada Gambar 1, yang memperlihatkan bagaimana data diorganisasi ke dalam tiga cluster secara sistematis.

W005A/T00 (W005A 1.5 G CVT WHITE_Matic	51	5	4	3
G0321M/T00 (H ACE PREMIO 2.8 M/T) WHITE_Manual	51	10	6	3
W005A/T00 (W005A 1.5 G CVT WHITE_Matic	51	5	4	3
W010A/T00 (W010A 1.5 G CVT WHITE_Matic	51	5	4	3
A2151GA/T00 (RAOZ 1.2 G CVT ONE TONE) WHITE_Matic	51	5	4	3
F005A/T00 (RUSH 1.5 A/T GR SPORT) WHITE_Matic	51	5	4	3
F005A/T00 (RUSH 1.5 A/T GR SPORT) WHITE_Matic	51	5	4	3
MAGT00GA/T00 (BINOVA ZENIX HYBRID 2.0 G HV CVT NPIC) DARK STEEL_Matic	45	5	6	3
Jumlah Cluster				
Cluster	Jumlah			
2	889			
1	474			
3	138			

Gambar 1. Hasil Perhitungan sistem K-Means

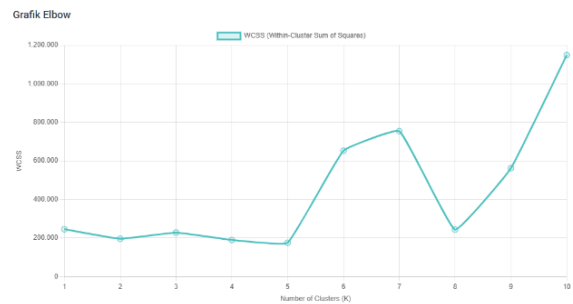
Pada Gambar 2, diagram scatter digunakan untuk memberikan visualisasi pola distribusi kendaraan dalam tiga cluster tersebut. Cluster 1, yang ditandai dengan warna merah, mencakup 474 kendaraan. Cluster ini memperlihatkan karakteristik dominan dalam hal preferensi terhadap warna tertentu dan kecenderungan menggunakan transmisi mesin dengan tingkat keunggulan yang tinggi. Cluster 2, yang ditandai dengan warna hijau dan mencakup 889 kendaraan, menunjukkan preferensi pelanggan yang konsisten terhadap warna tertentu dan pilihan transmisi yang relatif seragam. Sedangkan Cluster 3, yang ditandai dengan warna biru dan mencakup 138 kendaraan, memperlihatkan pola preferensi warna yang lebih bervariasi, namun dengan kecenderungan memilih jenis transmisi mesin yang lebih rendah dibandingkan cluster lainnya. Distribusi ini tidak hanya memberikan informasi mengenai karakteristik utama masing-masing cluster, tetapi juga menjadi dasar untuk memahami pola minat pelanggan secara lebih mendalam.



Gambar 2. Diagram Scatter

Penentuan jumlah cluster yang optimal dilakukan dengan menggunakan Metode Elbow, sebagaimana disajikan pada Gambar 3. Grafik Metode Elbow ini menampilkan hubungan antara jumlah cluster (k) dengan nilai *inertia*, yaitu ukuran variansi data dalam cluster. Berdasarkan grafik tersebut, terlihat bahwa titik ideal untuk jumlah cluster adalah ketika nilai $k=3$, di mana grafik mulai mendatar. Hal ini menunjukkan bahwa penambahan cluster lebih lanjut tidak memberikan pengurangan signifikan pada variansi, sehingga $k=3$ dianggap sebagai pilihan optimal untuk pengelompokan. Proses evaluasi ini penting untuk memastikan bahwa hasil pengelompokan yang

dihasilkan tidak hanya relevan, tetapi juga efisien dalam menangkap pola data yang ada.



Gambar 3. Elbow

Melalui proses pengelompokan ini, Auto2000 Salemba dapat memperoleh wawasan yang lebih mendalam tentang pola preferensi pelanggan, yang dapat digunakan untuk mengoptimalkan strategi pemasaran dan penyesuaian penawaran produk agar lebih sesuai dengan kebutuhan pasar.

5. KESIMPULAN DAN SARAN

Menerapkan metode K-Means, penelitian ini berhasil mengelompokkan data penjualan mobil Auto2000 Salemba ke dalam tiga klaster berdasarkan atribut varian, warna, dan harga. Klaster 1 terdiri dari 474 kendaraan, Klaster 2 mencakup 889 kendaraan, dan Klaster 3 berisi 138 kendaraan. Setiap klaster merepresentasikan preferensi pelanggan yang berbeda. Hasil pengelompokan ini memberikan wawasan strategis untuk pengembangan strategi pemasaran, seperti fokus pada promosi model di klaster dengan penjualan tinggi serta memastikan ketersediaan varian yang populer. Penelitian ini memberikan dasar bagi Auto2000 untuk memahami pola penjualan lebih baik dan mengidentifikasi tren pasar di masa mendatang. Pada penelitian berikutnya, disarankan untuk memperluas analisis dengan memasukkan variabel tambahan seperti data demografi pelanggan.

DAFTAR PUSTAKA

- [1] Annur, H., 2019. Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering. *Jurnal Informasi Upgris*, 5(1).
- [2] Fajrin, D.R., Triayudi, A., & Ningsih, S., 2022. Analisis Faktor yang mempengaruhi Pembelian Makanan secara Online pada Masa Pandemi Covid-19 menggunakan Metode K Means Clustering (Studi Kasus: Online Shop Bellyboys.id). *Jurnal Teknologi Informasi dan Komunikasi*, 6(1), p.77.
- [3] Li, K., et al., 2019. Impact factors analysis on the probability characterized effects of time of use demand response tariffs using association rule mining method.
- [4] Ilyas Sahputra, M., Triayudi, A., & Rubhasy, A., 2022. Analisis Faktor Yang Mempengaruhi Penumpang Angkutan Umum Beralih ke

- Transportasi Online Go-Jek menggunakan Metode K-Means Clustering. *Jurnal Teknologi Informasi dan Komunikasi*, 6(1), p.63.
- [5] Oktarian, S., Defit, S., & Sumijan, 2020. Klasterisasi Penentuan Minat Siswa dalam Pemilihan Sekolah Menggunakan Metode Algoritma K-Means Clustering. *Jurnal TEKNO KOMPAK*, 15(2).
- [6] Reynaldo, Y., Triayudi, A., & Ningsih, S., 2022. Analisis Faktor yang Mempengaruhi Gamers PC dan Konsol Beralih ke Game Mobile menggunakan Metode K-Means Clustering. *Jurnal Teknologi Informasi dan Komunikasi*, 6(1), p.42.
- [7] Wardhana, M., Triayudi, A., & Hayati, N., 2022. Analisis Faktor Calon Nasabah PT. Bank Central Asia dalam Pembuatan Rekening Online menggunakan Metode K-Means Clustering Studi Kasus Wisma Asia BCA. *Jurnal Teknologi Informasi dan Komunikasi*, 6(1), p.85.
- [8] Cholissodin, I., & Soebroto, A.A., 2019. *AI, Machine Learning & Deep Learning (Teori & Implementasi)*.
- [9] Maulia Usnaini, V.Y., 2021. Perancangan Sistem Informasi Pendataan Penduduk Berbasis Web Menggunakan Metode Waterfall. *Jurnal Manajemen Informatika Jayakarta*, 1, pp.36–55.
- [10] Utomo, A. N., & Alfaridzi, M., 2018. Perancangan Sistem Informasi Pada Percetakan CV Citra Kencana Jakarta Timur Berbasis Web. *Jurnal Rekayasa Informasi*, 7(1), 43–47.