PENERAPAN ALGORITMA K-MEANS UNTUK KLASTERISASI MOBIL BEKAS BERDASARKAN FAKTOR PENENTU HARGA JUAL

Daffa Arkan, Shofa Shofiah Hilabi, Elfina Novalia, Baenil Huda
Sistem Informasi, Universitas Buana Perjuangan Karawang
Jl. H.S. Ronggowaluyo Telukjambe Timur, Karawang-41361, Indonesia
si22.daffaarkan@mhs.ubpkarawang.ac.id

ABSTRAK

Penentuan harga jual pada pasar mobil bekas memiliki kompleksitas tinggi karena dipengaruhi oleh berbagai faktor teknis yang dinamis. Faktor-faktor seperti tahun produksi dan jarak tempuh sering kali menimbulkan ambiguitas bagi konsumen dalam melakukan segmentasi kendaraan secara objektif. Penelitian ini bertujuan untuk mengimplementasikan algoritma K-Means Clustering dalam melakukan segmentasi harga mobil bekas guna memetakan kasta ekonomi kendaraan secara sistematis. Metodologi yang digunakan mencakup tahapan Knowledge Discovery in Database (KDD) terhadap 1.222 entri data, yang meliputi preprocessing, transformasi data, hingga penentuan klaster optimal menggunakan metode Elbow dan validasi Silhouette Score. Hasil penelitian menunjukkan bahwa jumlah klaster paling optimal adalah k=3 dengan tingkat kestabilan nilai Silhouette Score yang baik. Klasterisasi ini berhasil mengidentifikasi tiga segmen pasar utama: kelompok ekonomi (low-end) dengan rata-rata harga Rp171 juta, kelompok menengah (mid-range) sebesar Rp393 juta, dan kelompok eksklusif (high-end) yang mencapai Rp760 juta. Implementasi ini membuktikan efektivitas algoritma K-Means dalam menyediakan instrumen pendukung keputusan bagi pelaku bisnis dan konsumen untuk menilai kewajaran harga pasar secara akurat.

Kata kunci : *Data Mining, K-Means, Klasterisasi, Mobil Bekas, Segmentasi Pasar.*

1. PENDAHULUAN

Perkembangan pesat industri otomotif, khususnya pada pasar mobil bekas di Indonesia, mendorong kebutuhan akan analisis data yang lebih akurat dalam mendukung pengambilan keputusan baik bagi konsumen maupun pelaku bisnis [1]. Meningkatnya volume transaksi mobil bekas menimbulkan tantangan dalam menentukan harga jual yang sesuai dengan kondisi dan spesifikasi kendaraan. Harga mobil bekas dipengaruhi oleh berbagai faktor seperti merek, tahun produksi, kapasitas mesin, serta kondisi fisik kendaraan yang bersifat kompleks dan sulit dianalisis secara manual [2].

Kompleksitas tersebut menunjukkan bahwa penentuan harga mobil bekas tidak dapat dilakukan secara subjektif tanpa dukungan analisis data yang sistematis. Pemanfaatan data analitik dalam skala besar telah terbukti memberikan kontribusi signifikan terhadap efektivitas pengambilan keputusan berbasis data [3]. Oleh karena itu, diperlukan suatu pendekatan yang mampu mengolah berbagai variabel kendaraan secara simultan untuk menghasilkan segmentasi yang lebih objektif dan terukur.

Dalam konteks ini, teknik klasterisasi dalam data mining dapat digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan atribut [4]. Berbagai penelitian menunjukkan bahwa penerapan algoritma K-Means efektif dalam mengelompokkan data berdasarkan karakteristik numerik serta mendukung proses analisis dan pengambilan keputusan secara lebih sistematis [5]. Selain itu, metode ini juga telah terbukti mampu menghasilkan segmentasi pasar yang lebih jelas serta meningkatkan akurasi dalam analisis data kendaraan bekas [6], [7], [8].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menerapkan algoritma K-Means dalam mengelompokkan data mobil bekas berdasarkan faktor-faktor penentu harga jual. Hasil dari penelitian ini diharapkan dapat memberikan gambaran segmentasi pasar yang lebih jelas serta menjadi alat bantu dalam pengambilan keputusan yang lebih efisien dan berbasis data [9].

2. TINJAUAN PUSTAKA

2.1. Mobil Bekas dan Harga Jual

Pasar mobil bekas di Indonesia menunjukkan pertumbuhan signifikan dengan volume penjualan mencapai 1,8 juta unit pada tahun 2024, melampaui penjualan mobil baru akibat kenaikan harga unit baru yang tidak sebanding dengan pendapatan masyarakat [10]. Kompleksitas harga jual kendaraan dipengaruhi oleh berbagai faktor teknis dan nonteknis, di mana variabel tahun produksi, jarak tempuh, transmisi, dan fitur keselamatan menjadi prediktor utama [11]. Mengingat tingginya variasi karakteristik kendaraan yang menyebabkan rentang harga menjadi tidak seragam, diperlukan pendekatan analitis yang sistematis untuk mengolah variabel-variabel tersebut secara simultan, sehingga dapat dihasilkan pemetaan pola harga yang objektif guna mendukung proses pengambilan keputusan bagi penjual maupun pembeli."

2.2. Data Mining

Data mining merupakan proses ekstraksi pola atau informasi bermakna dari dataset berskala besar melalui integrasi metode statistik, *machine learning*, dan sistem basis data menggunakan metodologi

CRISP-DM[12]. Dalam konteks pasar mobil bekas, penerapan data mining melalui algoritma *K-Means* terbukti efektif untuk membentuk segmentasi pasar yang terukur berdasarkan kemiripan atribut numerik seperti harga, jarak tempuh, dan tahun produksi[13]. Melalui tahapan pengolahan data yang terstruktur mulai dari pembersihan hingga pemodelan pendekatan ini mampu mengungkap hubungan tersembunyi antar variabel spesifikasi kendaraan, sehingga memberikan landasan analitis yang kuat dan sistematis dalam mendukung pengambilan keputusan terkait penetapan harga jual maupun pemilihan kendaraan.

2.3. Klasterisasi

Klasterisasi merupakan teknik *data mining* yang bertujuan mengelompokkan objek ke dalam beberapa grup berdasarkan tingkat kemiripan yang diukur menggunakan fungsi jarak, seperti *Euclidean Distance* untuk data numerik[14]. Dalam industri otomotif, metode ini terbukti efektif untuk segmentasi pasar dan analisis pola penjualan dengan cara meminimalkan variasi di dalam klaster serta memaksimalkan perbedaan antar kelompok[15]. Penerapan klasterisasi pada data mobil bekas melalui proses pengolahan yang terstruktur memungkinkan terciptanya segmentasi terukur berdasarkan atribut harga, jarak tempuh, dan tahun produksi, yang pada akhirnya menyediakan informasi sistematis untuk mendukung pengambilan keputusan objektif bagi pelaku pasar.

2.4. Algoritma K-Means

Algoritma *K-Means* merupakan metode *unsupervised learning* yang berfungsi mengelompokkan data ke dalam k klaster berdasarkan kemiripan fitur melalui proses iteratif[16]. Mekanisme algoritma ini dimulai dengan inisialisasi *centroid* acak, pengelompokkan data berdasarkan jarak terdekat, hingga penghitungan ulang posisi *centroid* berdasarkan rata-rata anggota klaster sampai mencapai konvergensi[17]. Dalam menentukan kedekatan antar data, *K-Means* menggunakan perhitungan jarak *Euclidean* yang sangat efektif untuk menangkap variasi atribut numerik pada mobil bekas seperti harga, tahun produksi, dan jarak tempuh[18]. Optimalisasi hasil segmentasi ini didukung dengan penggunaan metode *Elbow* untuk menentukan jumlah klaster terbaik, sehingga mampu menghasilkan pemetaan kelompok kendaraan dengan karakteristik yang stabil dan representatif[19].

$$d(x, y) = \sqrt{\Sigma(x_i - y_i)}^2 \quad (1)$$

Pengelompokan data mobil bekas dilakukan menggunakan algoritma K-Means dengan perhitungan kemiripan berdasarkan jarak *Euclidean Distance* terhadap variabel harga, tahun, dan kilometer yang telah di normalisasi. Data dialokasikan ke dalam klaster dengan jarak minimum terdekat dari pusat klaster (*centroid*), yang kemudian kualitas strukturnya diuji menggunakan *Silhouette Coefficient* untuk mengukur seberapa optimal pemisahan antar kelompok yang terbentuk. Hasil evaluasi yang mendekati nilai 1 mengindikasikan bahwa segmentasi harga telah terbagi secara konsisten dengan struktur klaster yang kuat dan minim tumpang tindih (*overlap*)

2.5. Google Colab

Google Colab merupakan platform berbasis *cloud* yang menyediakan lingkungan *Jupyter Notebook* untuk keperluan komputasi data tanpa memerlukan instalasi perangkat lunak secara lokal. Platform ini menawarkan akses gratis terhadap sumber daya komputasi tinggi seperti GPU dan TPU, yang sangat krusial dalam pemrosesan dataset besar dan eksperimen machine learning. Dalam penelitian data *science*, Google Colab terbukti efektif untuk melakukan web scraping, pemrosesan data, hingga visualisasi secara komprehensif tanpa hambatan spesifikasi perangkat keras. Penggunaan Colab dalam penelitian ini memungkinkan seluruh tahapan data mining dan klasterisasi berjalan secara efisien, fleksibel, serta mendukung aspek reproducibility pada analisis data pasar mobil bekas. Selain itu, fitur integrasi langsung dengan berbagai pustaka Python populer seperti Pandas dan *Scikit-Learn* sangat memudahkan proses sinkronisasi data secara real-time. Kemampuan penyimpanan berbasis cloud juga memastikan bahwa seluruh dokumentasi eksperimen tersimpan secara aman dan dapat diakses kembali untuk pengembangan model di masa mendatang.

2.6. Python

Python merupakan bahasa pemrograman tingkat tinggi bersifat interpreted yang menjadi standar utama dalam riset machine learning dan analisis data. Keunggulan utamanya terletak pada sintaks yang sederhana serta modularitas tinggi, sehingga memudahkan penulisan dan pemeliharaan kode saat menangani dataset yang kompleks. Dalam ekosistem data mining, Python didukung oleh pustaka (library) yang lengkap untuk manipulasi data, komputasi numerik, hingga pemodelan algoritma dalam satu lingkungan kerja. Hal ini memungkinkan seluruh proses, mulai dari pra-pengolahan data hingga evaluasi model, berjalan secara efisien dan dapat direplikasi (reproducible) tanpa memerlukan integrasi bahasa pemrograman lain.

2.7. Penelitian Terkait

Sejumlah penelitian terdahulu telah menerapkan algoritma K-Means untuk berbagai kebutuhan segmentasi data[20]. Melakukan pemodelan keputusan konsumen pada pembelian mobil bekas dengan membagi data ke dalam tiga klaster berdasarkan variabel merek, harga, dan tahun produksi. Implementasi ini didasarkan pada fakta bahwa segmentasi awal menggunakan K-Means mampu meningkatkan akurasi model prediksi harga dibandingkan dengan pengujian tanpa klasterisasi. Selain itu, penerapan K-Means juga terbukti efektif

dalam memetakan unit kendaraan berdasarkan variabel harga dan jarak tempuh guna mendukung optimalisasi manajemen armada secara lebih sistematis.

Di luar sektor otomotif secara umum, identifikasi tipe kendaraan terlaris melalui teknik data mining terbukti efektif untuk merumuskan strategi penjualan yang lebih terukur dan berbasis data[21]. Pemanfaatan algoritma K-Means pada sistem rekomendasi mobil bekas juga dapat didukung dengan teknik Principal Component Analysis (PCA) untuk meningkatkan kejelasan visualisasi pada pusat klaster Integrasi kedua metode ini memungkinkan identifikasi struktur kelompok data yang lebih presisi melalui reduksi dimensi tanpa menghilangkan informasi esensial dari variabel penelitian. Fleksibilitas algoritma ini juga terbukti dalam pengelompokan kinerja UKM ke dalam tiga tingkatan prioritas pembinaan [15]. Secara kolektif, berbagai studi tersebut mengonfirmasi bahwa K-Means merupakan metode yang konsisten dan efektif dalam menangkap variasi pola data numerik melalui perhitungan jarak Euclidean. Kemampuan algoritma ini dalam melakukan segmentasi berbasis kedekatan jarak antardata menjadikannya instrumen yang relevan untuk berbagai domain penelitian, termasuk dalam analisis pasar kendaraan bekas.

3. METODE PENELITIAN

3.1. Objek Penelitian

Objek dalam penelitian ini adalah dataset mobil bekas yang beredar di pasar otomotif Indonesia. Data tersebut mencakup berbagai atribut teknis dan harga jual yang merepresentasikan kondisi pasar aktual. Dataset yang digunakan merupakan gabungan dari sumber data sekunder (Kaggle) dan data primer yang diperoleh melalui proses web scraping pada platform jual beli daring (OLX). Total data yang dianalisis berjumlah 1.122 baris, yang memuat fitur-fitur utama seperti merek, tahun produksi, jarak tempuh (mileage), kapasitas mesin, jenis transmisi, serta harga jual.

Penelitian ini dikategorikan sebagai non-field research karena seluruh proses dilakukan melalui analisis data digital tanpa observasi lapangan secara fisik. Fokus utama analisis adalah mengidentifikasi pola kemiripan atribut numerik kendaraan menggunakan algoritma K-Means untuk membentuk segmentasi harga yang objektif.

Seluruh tahapan pengolahan data, mulai dari data *cleaning*, transformasi, normalisasi, hingga pengujian *Silhouette Score*, dilakukan menggunakan bahasa pemrograman Python. Lingkungan komputasi yang digunakan adalah Google Colab, yang dipilih karena kemampuannya dalam menangani pemrosesan data besar secara efisien serta mendukung visualisasi hasil klasterisasi secara menyeluruh.

3.2. Sampel Dataset

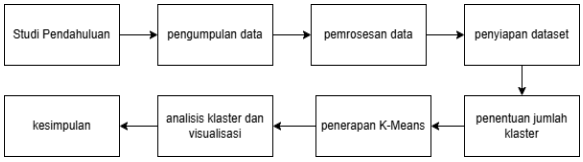
Dataset yang digunakan dalam penelitian ini merupakan data mobil bekas yang diperoleh dari berbagai sumber penjualan kendaraan, termasuk data publik dan hasil scraping pada platform jual beli daring. Data yang dikumpulkan memiliki beberapa atribut utama, antara lain merek, model, tahun produksi, jarak tempuh, kapasitas mesin, jenis transmisi, dan harga. Dataset ini digunakan sebagai dasar dalam proses pengolahan data dan analisis menggunakan metode *K-Means* untuk menghasilkan pengelompokan mobil bekas berdasarkan karakteristik tertentu. Sampel dari 1.122 data yang digunakan ditampilkan pada Tabel 1 berikut.

Tabel 1. Sample Data

No	Nama Kendaraan	Tahun	Harga Jual (Rp)
1	AYLA X 1.2	2018	101.000.000
2	AGYA TRD 1.0	2015	82.000.000
3	X-TRAIL 2.5	2015	169.000.000
4	YARIS S TRD	2020	218.000.000
5	AGYA G 1.2	2019	117.000.000
6	INNOVA G 2.0	2017	285.000.000
7	FORTUNER VRZ	2022	540.000.000
8	BRIO SATYA	2019	145.000.000
9	XPANDER ULTIMATE	2021	255.000.000
10	JAZZ RS	2016	195.000.000

3.3. Prosedur Penelitian

Prosedur penelitian ini disusun secara sistematis untuk memastikan proses analisis data mobil bekas dapat dilakukan secara terukur dan objektif. Alur penelitian dimulai dari identifikasi kebutuhan data hingga tahap interpretasi hasil klasterisasi. Secara visual, alur penelitian ini digambarkan melalui diagram alir pada Gambar 1.



Gambar 1. Alur Penelitian

Prosedur penelitian ini disusun dalam empat tahapan utama yang saling berkaitan untuk memastikan proses analisis segmentasi harga mobil bekas dilakukan secara sistematis. Tahap pertama dimulai dengan studi pendahuluan dan perancangan kebutuhan data, di mana dilakukan penelusuran literatur terkait pasar otomotif dan konsep data mining. Pada fase ini, atribut kunci seperti merek, tahun produksi, jarak tempuh, dan harga jual ditentukan sebagai variabel utama penelitian, dibarengi dengan penyusunan kerangka alur kerja dalam bentuk diagram alir.

Tahap kedua berfokus pada pengumpulan dan prapemrosesan data, di mana 1.122 data dikompilasi dari berbagai platform jual beli daring dan dataset publik. Proses prapemrosesan dilakukan melalui identifikasi serta penghapusan data duplikat, penanganan *missing values*, serta transformasi atribut kategorikal menjadi numerik. Selain itu, dilakukan

normalisasi menggunakan *feature scaling* agar atribut dengan rentang nilai besar, seperti harga dan jarak tempuh, berada pada skala yang sebanding sebelum diproses oleh algoritma.

Tahap ketiga merupakan implementasi algoritma *K-Means* menggunakan bahasa pemrograman Python pada lingkungan Google Colab. Proses ini diawali dengan memuat dataset hasil prapemrosesan, dilanjutkan dengan penentuan jumlah klaster K optimal melalui metode *Elbow* dan *Silhouette Coefficient*. Algoritma kemudian bekerja secara iteratif mulai dari inisialisasi centroid, perhitungan jarak *Euclidean*, hingga pembaruan posisi pusat klaster sampai mencapai kondisi konvergen, di mana setiap data mobil bekas akhirnya memperoleh label klaster berdasarkan pola kemiripan atributnya.

Tahap terakhir adalah analisis dan pemanfaatan hasil klasterisasi, yang melibatkan pengamatan mendalam terhadap karakteristik unik pada setiap segmen harga yang terbentuk. Karakteristik tersebut ditinjau berdasarkan rata-rata atribut utama seperti usia kendaraan dan nilai jual, yang kemudian divisualisasikan menggunakan *scatter plot* untuk memberikan gambaran segmentasi pasar yang intuitif. Hasil akhir dari tahapan ini diharapkan dapat menjadi bahan pertimbangan strategis bagi penjual maupun calon pembeli dalam pengambilan keputusan berbasis data di pasar mobil bekas.

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Data

Dataset yang digunakan dalam penelitian ini merupakan hasil integrasi data dari berbagai sumber informasi kendaraan bekas secara daring. Data primer diperoleh melalui proses pengumpulan mandiri menggunakan teknik web scraping pada platform jual beli kendaraan, yang kemudian diperkaya dengan penggabungan dataset publik dari repositori data daring guna meningkatkan volume serta variasi data. Seluruh data yang dikumpulkan mencakup kendaraan dengan rentang tahun produksi yang variatif, mulai dari tahun 1990-an hingga 2024, sehingga mampu menangkap dinamika harga baik untuk mobil keluaran lama maupun model terbaru. Penggabungan kedua sumber data ini menghasilkan total populasi awal sebanyak 1.122 baris data yang mencakup berbagai segmentasi kendaraan di pasar mobil bekas.

Secara struktural, dataset ini memuat atribut-atribut kunci yang merepresentasikan nilai teknis dan ekonomi dari sebuah kendaraan. Atribut pertama adalah identitas kendaraan yang memuat informasi mengenai merk dan tipe unit. Atribut kedua adalah tahun produksi kendaraan yang dikategorikan sebagai variabel numerik diskrit; variabel ini sangat fundamental karena menjadi indikator usia pakai serta tingkat depresiasi nilai fisik kendaraan. Atribut ketiga adalah harga jual unit dalam satuan mata uang Rupiah yang merupakan variabel numerik kontinu. Kombinasi dari variabel tahun dan harga ini menjadi fitur utama yang diolah dalam algoritma K-Means untuk memetakan kasta ekonomi pasar mobil bekas.

4.2. Pemrosesan Data

Tahap prapemrosesan data merupakan langkah krusial dalam siklus pengolahan data untuk menjamin kualitas dan validitas hasil klasterisasi. Data yang telah dihimpun melalui proses integrasi tidak dapat langsung diolah menggunakan algoritma *K-Means* karena adanya ketidakteraturan format dan perbedaan skala nilai antar variabel. Langkah pertama yang dilakukan adalah pembersihan data (*data cleaning*), di mana variabel harga yang sebelumnya bertipe teks dibersihkan dari simbol mata uang serta tanda baca pemisah ribuan, kemudian dikonversi ke dalam format numerik. Selain itu, dilakukan penanganan terhadap data yang tidak lengkap (*missing values*) dengan menghapus baris pada variabel tahun guna menjaga akurasi perhitungan jarak antar titik data.

Setelah data dipastikan bersih secara administratif, dilakukan tahap transformasi menggunakan metode *Z-Score Standardizer*. Langkah ini sangat fundamental mengingat variabel tahun produksi dan harga jual memiliki satuan serta rentang nilai yang sangat jauh berbeda. Melalui proses normalisasi ini, setiap variabel ditransformasi sedemikian rupa sehingga memiliki rata-rata nol dan standar deviasi satu. Hal ini dilakukan untuk menyamakan bobot pengaruh kedua variabel dalam perhitungan algoritma, sehingga tidak ada satu variabel yang mendominasi hasil pengelompokan hanya karena besaran nilai nominalnya. Hasil dari prapemrosesan ini adalah dataset yang telah ternormalisasi dan secara teknis siap untuk dianalisis pada tahap penentuan klaster.

Struktur data hasil prapemrosesan yang telah melewati tahap normalisasi disajikan secara sistematis pada Tabel 2 untuk menunjukkan perubahan nilai sebelum masuk ke tahap algoritma.

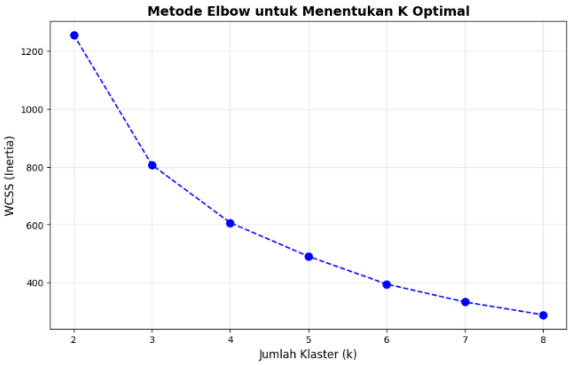
Tabel 2. Sampel Data Setelah Tahap Prapemrosesan

No	Tahun (Scaled)	Harga Jual (Scaled)
1	-0,090	-0,871
2	-0,946	-0,980
3	-0,946	-0,482
4	0,480	-0,202
5	0,195	-0,780
6	-0,376	0,183
7	1,050	1,642
8	0,195	-0,519
9	0,765	0,351
10	-0,661	-0,124

4.3. Penentuan Jumlah Klaster Optimal (*Elbow method*)

Tahap krusial dalam algoritma *K-Means* adalah menentukan jumlah klaster (k) yang paling optimal untuk merepresentasikan struktur data secara alami. Dalam penelitian ini, penentuan jumlah klaster dilakukan menggunakan Metode *Elbow*. Prinsip dasar dari metode ini adalah dengan menghitung nilai

Within-Cluster Sum of Squares (WCSS) atau inersia untuk berbagai nilai k. WCSS mengukur total jarak kuadrat antara setiap titik data dengan pusat klaster (*centroid*) masing-masing. Semakin kecil nilai WCSS, maka semakin rapat data dalam kelompoknya. Namun, penambahan jumlah klaster secara berlebihan akan mengakibatkan model menjadi tidak efisien (terjadi *overfitting*). Berdasarkan hasil pengujian yang dilakukan terhadap dataset mobil bekas, diperoleh grafik Elbow sebagaimana disajikan pada Gambar 2.

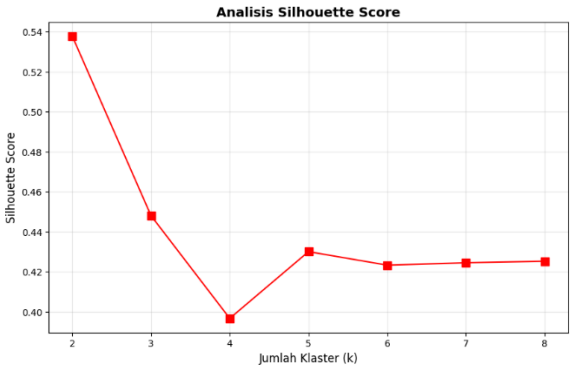


Gambar 2. Elbow Method

Berdasarkan visualisasi pada Gambar 2, terlihat bahwa sumbu horizontal (x) merepresentasikan jumlah klaster yang diuji, sedangkan sumbu vertikal (y) menunjukkan nilai inersia atau WCSS. Grafik menunjukkan terjadinya penurunan nilai WCSS yang sangat tajam pada saat jumlah klaster meningkat dari k=2 hingga k=3. Setelah melewati titik k=3, penurunan nilai WCSS mulai melandai dan tidak lagi menunjukkan perubahan yang signifikan secara drastis. Titik di mana terjadi perubahan sudut penurunan yang menyerupai bentuk siku (*elbow point*) inilah yang diidentifikasi sebagai jumlah klaster paling ideal. Berdasarkan hasil pengamatan tersebut, dapat disimpulkan bahwa k=3 merupakan jumlah klaster yang paling optimal untuk melakukan segmentasi pada dataset ini. Pemilihan jumlah klaster ini dianggap seimbang dalam memberikan hasil pengelompokan yang padat namun tetap mudah untuk diinterpretasikan secara sistematis.

4.4. Validasi Klaster (*Silhouette Score*)

Setelah mendapatkan indikasi jumlah klaster melalui metode *Elbow*, tahap selanjutnya adalah melakukan validasi menggunakan metode Silhouette Score. Metode ini bertujuan untuk mengukur seberapa baik sebuah titik data ditempatkan dalam klasternya dibandingkan dengan klaster lainnya. Nilai *Silhouette Score* berkisar antara -1 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa data telah terkelompok dengan sangat baik dan memiliki pemisahan yang jelas dari klaster tetangga. Sebaliknya, nilai yang mendekati 0 atau negatif mengindikasikan adanya tumpang tindih (*overlap*) antar kelompok. Hasil pengujian nilai koefisien Silhouette untuk berbagai jumlah klaster (k) disajikan pada Gambar 3.

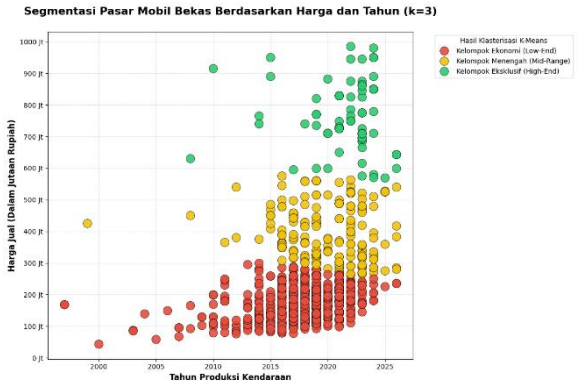


Gambar 3. Hasil Pengujian Silhouette Score

Berdasarkan grafik pada Gambar 3, dapat diamati bahwa nilai koefisien tertinggi diperoleh pada saat jumlah klaster k=2, yang kemudian diikuti oleh nilai pada k=3. Meskipun k=2 secara matematis memiliki skor tertinggi, penelitian ini memutuskan untuk menetapkan k=3 sebagai jumlah klaster final. Keputusan ini diambil berdasarkan pertimbangan kebutuhan analisis segmentasi pasar yang lebih mendalam, di mana pembagian menjadi tiga kelompok memberikan gambaran yang lebih representatif mengenai kasta ekonomi kendaraan (ekonomi, menengah, dan eksklusif). Selain itu, pemilihan k=3 juga bertujuan untuk menghindari pengelompokan yang terlalu umum (*under-fitting*) sehingga variasi data dapat tertangkap secara lebih spesifik. Nilai *Silhouette Score* pada k=3 menunjukkan angka yang positif dan cukup stabil, yang mengonfirmasi bahwa setiap klaster memiliki struktur yang solid dengan pemisahan antar kelompok yang dapat dipertanggungjawabkan secara ilmiah. Dengan demikian, hasil validasi ini memperkuat keputusan penggunaan tiga klaster untuk tahap analisis dan pembahasan selanjutnya.

4.5. Analisis Hasil Klasterisasi

Setelah menetapkan k=3 sebagai jumlah klaster optimal, algoritma *K-Means* dijalankan untuk mengelompokkan dataset mobil bekas berdasarkan variabel tahun produksi dan harga jual. Untuk meningkatkan akurasi pengelompokan, dilakukan pembobotan pada variabel harga guna memastikan segmentasi yang dihasilkan selaras dengan realitas ekonomi pasar. Hasil klasterisasi ini membagi data ke dalam tiga kategori formal, yaitu: Kelompok Ekonomi (*Low-End*), Kelompok Menengah (*Mid-Range*), dan Kelompok Eksklusif (*High-End*). Visualisasi distribusi data hasil klasterisasi secara lebih detail ditampilkan pada Gambar 4.

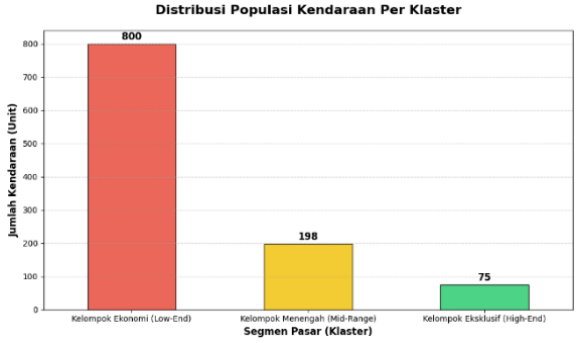


Gambar 4. Visualisasi Klasterisasi

Berdasarkan visualisasi pada Gambar 4, terlihat pemisahan yang jelas antar ketiga kelompok tersebut. Kelompok Ekonomi (*Low-End*) yang ditandai dengan warna merah mendominasi area harga rendah, mencakup kendaraan dengan tahun produksi yang lebih tua atau mobil kategori LCGC. Kelompok Menengah (*Mid-Range*) dengan warna kuning mengisi segmen pusat yang menawarkan keseimbangan antara harga kompetitif dan tahun produksi yang relatif muda. Sementara itu, Kelompok Eksklusif (*High-End*) yang ditandai dengan warna hijau menempati posisi puncak dengan karakteristik harga tertinggi dan tahun produksi terbaru. Implementasi pembobotan harga terbukti efektif dalam memisahkan kendaraan dengan nilai ekonomi tinggi agar tidak tercampur ke dalam kelompok ekonomi meskipun memiliki usia kendaraan yang serupa. Hasil ini memberikan gambaran komprehensif mengenai struktur kasta harga pada pasar mobil bekas secara sistematis.

4.6. Distribusi Populasi Tiap Klaster

Setelah proses klasterisasi selesai, dilakukan analisis terhadap jumlah anggota atau populasi dari masing-masing segmen pasar yang terbentuk. Analisis populasi ini penting untuk memberikan gambaran mengenai dominasi tipe kendaraan yang beredar di pasar mobil bekas berdasarkan dataset yang telah dihimpun. Distribusi jumlah kendaraan untuk setiap kelompok segmen pasar disajikan dalam bentuk diagram batang pada Gambar 5.

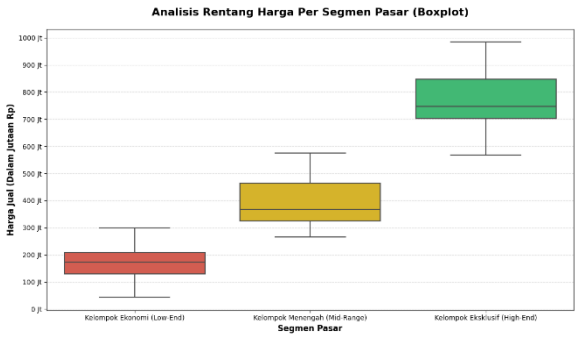


Gambar 5. Distribusi Jumlah Kendaraan Per Segmen Pasar

Berdasarkan Gambar 5, dapat dilihat bahwa Kelompok Ekonomi (*Low-End*) memiliki jumlah anggota yang paling dominan dibandingkan dengan kelompok lainnya. Hal ini menunjukkan bahwa ketersediaan unit mobil bekas di pasar didominasi oleh kendaraan dengan harga terjangkau atau unit dengan usia pemakaian yang lebih lama. Sementara itu, Kelompok Menengah (*Mid-Range*) memiliki populasi yang cukup signifikan sebagai penyeimbang pasar, dan Kelompok Eksklusif (*High-End*) menjadi segmen dengan jumlah unit paling sedikit. Struktur populasi ini membentuk pola piramida, di mana semakin tinggi kasta ekonomi suatu kendaraan, maka semakin selektif dan sedikit jumlah ketersediaan unitnya di pasar. Hasil ini mengonfirmasi bahwa permintaan dan penawaran terbesar masih berada pada segmen ekonomi dan menengah.

4.7. Analisis Rentang Harga Menggunakan Boxplot

Untuk memvalidasi konsistensi pengelompokan harga pada setiap klaster, dilakukan analisis menggunakan grafik *Boxplot*. Visualisasi ini bertujuan untuk melihat distribusi data, nilai tengah (*median*), serta rentang harga minimum dan maksimum pada setiap segmen pasar yang telah terbentuk. Melalui grafik ini, dapat diidentifikasi sejauh mana perbedaan kasta ekonomi antar klaster dan apakah terdapat tumpang tindih nilai harga yang signifikan. Visualisasi rentang harga per segmen pasar disajikan pada Gambar 6.



Gambar 6. Analisis Rentang Harga Menggunakan Boxplot

Berdasarkan visualisasi pada Gambar 6, dapat ditarik beberapa kesimpulan mengenai struktur harga pada setiap kelompok. Pertama, Kelompok Ekonomi (*Low-End*) memiliki rentang harga yang paling sempit dan terkonsentrasi di nilai bawah, yang menunjukkan keseragaman harga pada unit kendaraan kategori ini. Kedua, Kelompok Menengah (*Mid-Range*) menunjukkan peningkatan rentang harga yang stabil dengan nilai median yang berada di posisi tengah, mencerminkan variasi unit yang lebih beragam. Ketiga, Kelompok Eksklusif (*High-End*) memiliki kotak sebaran yang paling tinggi dengan jangkauan harga yang luas, mengindikasikan bahwa pada segmen premium, nilai kendaraan sangat bervariasi tergantung

pada spesifikasi dan tahun produksi terbaru. Keberadaan beberapa titik di luar garis (*outliers*) menunjukkan adanya unit kendaraan dengan harga khusus yang tetap masuk ke dalam klaster tersebut karena kemiripan karakteristik tahun produksinya. Secara keseluruhan, grafik *Boxplot* ini mengonfirmasi bahwa setiap klaster memiliki batas harga yang terpisah secara jelas.

4.8. Profil dan Karakteristik Tiap Klaster

Tahap akhir dari analisis hasil adalah mengidentifikasi profil unik dari setiap klaster yang telah terbentuk. Proses *profiling* dilakukan dengan menghitung nilai rata-rata (*mean*) dari variabel tahun produksi dan harga jual pada masing-masing kelompok. Hal ini bertujuan untuk memberikan label yang representatif bagi tiap klaster sesuai dengan karakteristik dominan anggota di dalamnya. Ringkasan karakteristik nilai rata-rata dari ketiga klaster disajikan pada Tabel berikut.

Tabel 3. Profil Karakteristik Klaster

Segmen Pasar	Rata-rata Tahun	Rata-rata Harga (Rp)	Populasi
Kelompok Ekonomi (Low-End)	2018	171.043.064	800 unit
Kelompok Menengah (Mid-Range)	2020	393.320.167	198 unit
Kelompok Eksklusif (High-End)	2021	760.893.307	75 unit

Berdasarkan Tabel tersebut, karakteristik dari masing-masing segmen dapat dijelaskan secara mendalam. Kelompok Ekonomi (*Low-End*) merupakan segmen terbesar dengan populasi 800 unit, memiliki rata-rata tahun produksi 2018 dan harga paling terjangkau di kisaran Rp 171 juta. Segmen ini didominasi oleh kendaraan fungsional dan mobil kota. Kelompok Menengah (*Mid-Range*) menunjukkan peningkatan kualitas dengan rata-rata tahun produksi yang lebih muda (2020) dan harga di kisaran Rp 393 juta. Sementara itu, Kelompok Eksklusif (*High-End*) mewakili kasta tertinggi dalam dataset dengan rata-rata tahun produksi terbaru (2021) dan nilai ekonomi mencapai Rp 760 juta.

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa algoritma *K-Means* sangat efektif dalam melakukan segmentasi pasar mobil bekas ke dalam tiga kasta ekonomi yang jelas, yaitu ekonomi (*low-end*), menengah (*mid-range*), dan eksklusif (*high-end*), dengan dukungan validasi metode *Elbow* dan *Silhouette Score* yang menunjukkan k=3 sebagai jumlah klaster paling optimal. Pembagian ini terbukti mampu memberikan gambaran objektif mengenai kewajaran harga jual berdasarkan variabel tahun produksi, harga, dan jarak tempuh kendaraan. Adapun saran untuk penelitian selanjutnya adalah menambahkan variabel yang lebih spesifik seperti kondisi fisik atau riwayat servis kendaraan, serta mencoba membandingkan hasil ini dengan algoritma klasterisasi lain seperti *K-Medoids* untuk mendapatkan tingkat akurasi yang lebih mendalam dan komprehensif.

DAFTAR PUSTAKA

[1] E. Rahmawati and P. Maulidya Effendi, "Pemodelan keputusan konsumen untuk pembelian mobil bekas menggunakan K-Means Clustering," *INFOTECH : Jurnal Informatika & Teknologi*, vol. 6, no. 1, pp. 22–33, Jun. 2025, doi: 10.37373/infotech.v6i1.1587.

[2] W. Maesaroh, T. M. Diansyah, R. Liza, Y. Fitri, and A. Lubis, "Pemanfaatan Algoritma K-Means Clustering Pada Sistem Rental Mobil," *Journal of Informatics Management and Information Technology*, vol. 5, no. 3, pp. 406–413, 2025, doi: 10.47065/jimat.v5i3.391.

[3] S. Shofiah Hilabi *et al.*, "Pemanfaatan Data Analitik dalam Big Data: Studi Kasus Implementasi di Pemerintahan," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 12, no. 1, pp. 378–390, 2025, [Online]. Available: http://jurnal.mdp.ac.id

[4] Awaljan, Tukino, Elfina Novalia, and Sandi Ahmad, "KLASIFIKASI HASIL PENJUALAN MINUMAN RINGAN PADA KOPERASI BERDASARKAN JENIS BARANG MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," Dec. 2022.

[5] A. Lia Hananto *et al.*, "Analysis of Drug Data Mining with Clustering Technique Using K-Means Algorithm," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jul. 2021. doi: 10.1088/1742-6596/1908/1/012024.

[6] K. Kasini, H. Rusnedy, L. S. Tanjung, and N. Y. S. Munti, "Penerapan Algoritma K-Means untuk Pengelompokan Data Mahasiswa Baru Program Studi Teknik Informatika di Universitas Pahlawan Tuanku Tambusai," *Jurnal Teknik Industri Terintegrasi*, vol. 8, no. 1, pp. 1378–1387, Jan. 2025, doi: 10.31004/jutin.v8i1.41449.

[7] G. R. Ramadhan, A. B. Nayla, and S. Riatma, "Used Car Price Prediction Using K-Means Clustering and Comparison of Four Supervised Learning Regression Algorithms," *ROUTERS: Jurnal Sistem dan Teknologi Informasi*, pp. 94–103, Jul. 2025, doi: 10.25181/rt.v3i2.4249.

[8] N. K. Juliani, N. M. S. Iswari, and N. W. Utami, "PENERAPAN DATA MINING UNTUK MENENTUKAN KELAYAKAN KENDARAAN SEPEDA MOTOR BEKAS MENGGUNAKAN ALGORITMA C4.5," *INFOTECH journal*, vol. 11, no. 2, pp. 234–241, Aug. 2025, doi: 10.31949/infotech.v11i2.15338.

[9] R. E. Adinagoro, I. Dwi, R. Premana, R. B. Pamungkas, P. Aryasetya, and D. Joedhistiro,

"Pengembangan Sistem Prediksi Harga Mobil Bekas Di Pasar India Menggunakan Algoritma XGBoost dan NLP," p. 2025.

[10] C. Whenjaya, H. Susanto, and Herman, "Model Prediksi Harga Mobil Bekas Menggunakan Regresi Linear Berganda Berdasarkan Fitur Kendaraan," *Journal of Digital Ecosystem for Natural Sustainability*, vol. 4, no. 2, pp. 95–99, Dec. 2024, doi: 10.63643/jodens.v4i2.263.

[11] A. Saepuloh and handoko Bambang, "PENGARUH KUALITAS PRODUK, HARGA, DAN PELAYANAN TERHADAP KEPUTUSAN PEMBELIAN MOBIL BEKAS PADA SHOWROOM MOBIL NURAYHAN MOTOR," *Jurnal Masharif al-Syariah: Jurnal Ekonomi dan Perbankan Syariah*, vol. 10, 2025.

[12] A. Agustino Maulana, K. Latifah, and N. D. Saputro, "IMPLEMENTASI ALGORITMA RANDOM FOREST REGRESSIONUNTUK ESTIMASI HARGA MOBIL BEKAS MEREK TOYOTA," *Jurnal Informatika Teknologi dan Sains*.

[13] I. Ependi and A. Purnama, "Implementasi Algoritma K-Means Clustering Untuk Pemilihan Dan Rekomendasi Mobil Bekas," *JURNAL INFORMATIKA*, vol. 11, no. 1, pp. 1–8, 2024, [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/ji

[14] M. A. Wardana and S. Suherman, "Penerapan Data Mining Pengelompokan Data Penjualan Motor di PT. Nusantara Surya Sakti Soppeng Menggunakan Metode K-Means Clustering," *Jurnal Ilmiah Sistem Informasi dan Teknik Informatika (JISTI)*, vol. 8, no. 1, pp. 123–132, Apr. 2025, doi: 10.57093/jisti.v8i1.282.

[15] S. Nurhalimah, S. Wahyuni, and W. Handoko, "Pengklasifikasian Tingkat Penjualan Sparepart Mobil di Putra Motor Menggunakan Algoritma K-Means Clustering," *Fusion : Journal of Research in Engineering*, vol. 2, 2025.

[16] Henry Adam, Tukino, Elfina Novalia, and A. L. Hananto, "Prediksi Penjualan Barang Menggunakan Metode K-Means dan Regresi Linear," *Bulletin of Computer Science Research*, vol. 5, no. 4, pp. 298–307, Jun. 2025, doi: 10.47065/bulletincsr.v5i4.541.

[17] B. A. Hendriansyah, A. Tri, J. Harjanta, and K. Latifah, "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING PADA SISTEM INFORMASI GEOGRAFIS FASILITAS KESEHATAN BPJS KESEHATAN KOTA SEMARANG," 2025.

[18] I. Ferdiansyah, B. Huda, A. Hananto, and U. Buana Perjuangan Karawang, "Analisis Clustering Menggunakan Metode K-Means Pada Kemiskinan Di Jawa Timur Tahun 2020," 2024.

[19] S. A. Fahmi, A. Ikhwan, and F. H. Sibarani, "Implementasi Data Mining Dengan Menggunakan Metode K-Means Clustering Untuk Menentukan Penjualan Sparepart Mobil," *Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI)*, vol. 4, no. 1, pp. 204–209, Jun. 2025, doi: 10.62712/juktisi.v4i1.388.

[20] E. Rahmawati and P. Maulidya Effendi, "Pemodelan keputusan konsumen untuk pembelian mobil bekas menggunakan K-Means Clustering," *INFOTECH : Jurnal Informatika & Teknologi*, vol. 6, no. 1, pp. 22–33, Jun. 2025, doi: 10.37373/infotech.v6i1.1587.

[21] M. L. Rifky, Z. Nugraha, M. B. Saputra, D. Pratama, E. Raswir, and Y. Pratama, "Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM) Implementasi Data Mining Untuk Penjualan Mobil Menggunakan Metode Naive Bayes." [Online]. Available: https://ejournal.unama.ac.id/index.php/jakakom