

SEGMENTASI PELANGGAN DAN PENARGETAN PEMBELI MENGGUNAKAN K-MEANS CLUSTERING: STUDI KASUS INDUSTRI REAL ESTATE**Razanah Rashif¹, Priza Pandunata*²**¹Program Studi Sistem Informasi, Universitas Jember, razhifrazanah@gmail.com²Program Studi Sistem Informasi, Universitas Jember, priza@unej.ac.id

*)Korespondensi: priza@unej.ac.id

Abstrak

Penelitian ini bertujuan untuk menganalisis segmentasi pelanggan dan penargetan pembeli dengan menggunakan metode K-Means Clustering pada industri real estate, dengan studi kasus PT Sembilan Bintang Lestari di Jember. Perusahaan ini menghadapi tantangan dalam memahami karakteristik pembeli secara spesifik, sehingga memerlukan pendekatan berbasis data untuk meningkatkan efektivitas pemasaran. Proses analisis dilakukan berdasarkan kerangka kerja CRISP-DM yang mencakup tahapan pemahaman bisnis, pemahaman data, persiapan data, pemodelan, dan evaluasi. Data pelanggan yang digunakan mencakup variabel usia, pendapatan, pekerjaan, tipe rumah, metode pembayaran, dan status pernikahan. Tahap pemodelan menggunakan algoritma K-Means dengan bantuan metode Elbow untuk menentukan jumlah kluster optimal, yang diperoleh sebanyak tiga kluster. Hasil segmentasi kemudian dievaluasi menggunakan Silhouette Score dan divisualisasikan melalui diagram scatter, boxplot, dan pie chart untuk memberikan gambaran komprehensif mengenai karakteristik masing-masing kluster. Hasil ini digunakan sebagai dasar untuk merancang strategi penargetan pembeli yang lebih spesifik dan tepat sasaran. Penelitian ini memberikan kontribusi dalam penerapan metode segmentasi berbasis data mining untuk mendukung pengambilan keputusan pemasaran secara lebih efektif di sektor real estate. Kebaruan penelitian ini terletak pada integrasi metode K-Means, Elbow, dan Silhouette Score secara sistematis pada data pelanggan real estate selama satu dekade, serta penyajian visualisasi segmentasi yang aplikatif untuk mendukung penargetan pembeli di tingkat operasional perusahaan.

Kata Kunci: K-Means Clustering, customer segmentation, real estate, Elbow method, data visualization.

Abstract

This study aims to analyze customer segmentation and targeting buyers using the K-Means Clustering method in the real estate industry, with the case study of PT Sembilan Bintang Lestari in Jember. The company faces challenges in understanding the buyer's specific characteristics, so that it requires a data -based approach to increase marketing effectiveness. The analysis process is carried out based on a crisp-DM framework which includes stages of business understanding, data understanding, data preparation, modeling, and evaluation. Customer data used includes age variables, income, jobs, house types, payment methods, and marital status. The modeling stage uses the K-Means algorithm with the help of the Elbow method to determine the optimal number of clusters, as many as three clusters. The segmentation results are then evaluated using the Silhouette Score and visualized through the scatter diagram, boxplot, and pie chart to provide a comprehensive picture of the characteristics of each cluster. This result is used as a basis for designing a more specific and targeted buyer targeting strategy. This study contributes to the application of data mining -based segmentation methods to support marketing decision making more effectively in the real estate sector. The novelty of this study lies in the integration of the K-Means, Elbow, and Silhouette Score methods in real estate customer data for a decade, as well as the presentation of applicative segmentation visualization to support buyers targeting at the company's operational level.

Keywords: K-Means Clustering, customer segmentation, real estate, Elbow method, data visualization.

I. PENDAHULUAN

Industri properti di Indonesia mengalami perkembangan pesat dalam beberapa tahun terakhir, seiring meningkatnya kebutuhan akan hunian dan properti komersial. Data dari Bank Indonesia (2024) mencatat bahwa pada triwulan pertama 2024, permintaan properti komersial meningkat sebesar 3,41% secara tahunan, naik signifikan dari 1,91% pada

triwulan keempat 2023 [1]. Pertumbuhan ini mencerminkan perbaikan kondisi ekonomi pascapandemi serta meningkatnya minat masyarakat terhadap kepemilikan properti. Namun demikian, tingginya permintaan juga diiringi dengan semakin ketatnya persaingan antar pelaku industri, baik dari sisi produk, harga, maupun pendekatan pemasaran. Perusahaan properti kini dituntut untuk mampu menyesuaikan strategi pemasaran secara cepat, presisi,

dan berbasis data agar tetap kompetitif di tengah pasar yang dinamis [2].

Di tengah kondisi ini, perusahaan properti dituntut untuk mampu beradaptasi dengan cepat dan mengoptimalkan strategi pemasaran yang lebih personal dan berbasis data [3]. Salah satu perusahaan yang menghadapi tantangan serupa adalah PT Sembilan Bintang Lestari, pengembang real estate di Kabupaten Jember yang telah mengelola proyek unggulan Rembulan Hill Residence sejak tahun 2014. Meskipun proyek tersebut telah menjangkau lebih dari 2.500 pelanggan, efektivitas promosi yang dilakukan masih belum optimal. Strategi pemasaran yang bersifat umum dan tidak berbasis segmentasi menyebabkan rendahnya konversi penjualan, meskipun perusahaan telah memiliki tim pemasaran yang cukup besar dan fasilitas pendukung yang lengkap [4]. Oleh karena itu, dibutuhkan pendekatan baru yang dapat membantu perusahaan memahami pelanggan secara lebih mendalam dan menyusun strategi pemasaran yang lebih terarah.

Salah satu pendekatan yang relevan untuk menyelesaikan permasalahan tersebut adalah penerapan segmentasi pelanggan berbasis data mining. Segmentasi pelanggan adalah proses pengelompokan pelanggan ke dalam kelompok-kelompok homogen berdasarkan karakteristik tertentu seperti usia, pendapatan, pekerjaan, atau preferensi pembelian [5]. Segmentasi ini bertujuan untuk mengidentifikasi kelompok pembeli potensial dan memungkinkan perusahaan menyusun strategi pemasaran yang lebih efektif dan efisien [6]. Beberapa studi terdahulu menunjukkan bahwa perusahaan yang menerapkan segmentasi pelanggan cenderung memiliki kinerja pemasaran dan penjualan yang lebih baik dibandingkan yang tidak [7], [8].

Data mining berperan penting dalam mendukung proses segmentasi dengan mengekstraksi informasi penting dari data mentah pelanggan. Data mining menggabungkan teknik statistik, machine learning, dan pengolahan data besar untuk menemukan pola tersembunyi yang dapat digunakan dalam pengambilan keputusan strategis [9]. Salah satu metodologi standar yang banyak digunakan dalam implementasi data mining adalah *CRISP-DM (Cross-Industry Standard Process for Data Mining)*, yang mencakup enam tahapan: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan implementasi. Metodologi ini telah terbukti efektif dalam berbagai konteks industri dan memberikan panduan sistematis agar proses analisis tetap terarah [10].

Dalam proses segmentasi pelanggan, salah satu algoritma yang paling populer digunakan adalah *K-Means Clustering*. *K-Means* bekerja dengan membagi data ke dalam sejumlah kluster berdasarkan tingkat kemiripan, sehingga data yang berada dalam satu

kluster memiliki karakteristik yang serupa [11]. Keunggulan dari algoritma ini terletak pada kesederhanaannya, efisiensinya dalam menangani data berukuran besar, dan kemampuannya menghasilkan kluster yang jelas terpisah secara geometris [12]. Namun, salah satu tantangan utama dalam penerapan *K-Means* adalah menentukan jumlah kluster optimal. Untuk mengatasi hal ini, digunakan metode *Elbow*, yang mengevaluasi nilai *Within-Cluster Sum of Squares (WCSS)* pada berbagai nilai *K*, lalu mengidentifikasi titik siku pada grafik sebagai jumlah kluster yang paling tepat [13], [14].

Beberapa studi sebelumnya menunjukkan bahwa penerapan algoritma *K-Means* dalam segmentasi pelanggan memberikan hasil yang signifikan dalam berbagai sektor industri. Penelitian oleh [15] dalam konteks pemasaran properti menunjukkan bahwa algoritma *K-Means* efektif untuk mengelompokkan pelanggan berdasarkan karakteristik demografis dan preferensi hunian. Dengan segmentasi yang lebih presisi, perusahaan dapat merancang kampanye pemasaran yang lebih tertarget, sehingga meningkatkan efisiensi promosi dan potensi konversi. Studi serupa dilakukan oleh [16] dalam industri perbankan, yang menunjukkan bahwa penerapan *K-Means* mampu meningkatkan akurasi dalam menentukan target pasar serta berdampak positif terhadap kepuasan dan loyalitas pelanggan.

Dalam hal pemilihan jumlah kluster, [17] menegaskan pentingnya penggunaan metode *Elbow* untuk menentukan nilai *K* yang optimal. Menurut mereka, identifikasi jumlah kluster yang tepat akan meningkatkan kualitas dan stabilitas hasil pengelompokan, terutama dalam konteks data evaluasi multidimensi. Selain itu, penelitian oleh [18] juga memperkuat bahwa validasi hasil clustering menggunakan kombinasi *Elbow Method* dan *Silhouette Score* dapat meningkatkan performa segmentasi lebih dari 20%, khususnya pada data pelanggan real estate di Indonesia.

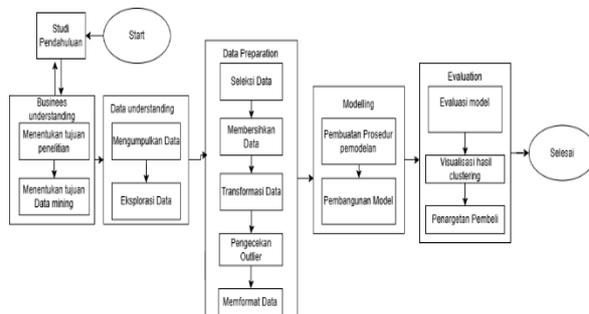
Menurut [5] menekankan pentingnya visualisasi data dalam proses analisis segmentasi. Mereka merekomendasikan penggunaan berbagai bentuk visualisasi seperti scatter plot, bar chart, dan pie chart untuk membantu tim pemasaran memahami profil setiap kluster secara intuitif dan cepat. Pendekatan ini terbukti mempermudah penyusunan strategi penargetan yang lebih efektif dan berbasis data aktual. Dengan demikian, integrasi antara algoritma *K-Means*, metode *Elbow*, dan visualisasi multivariabel menjadi strategi menyeluruh yang terbukti mendukung keputusan pemasaran di berbagai sektor industri, termasuk properti.

Penelitian ini bertujuan untuk menganalisis dan mengidentifikasi segmen pelanggan potensial PT Sembilan Bintang Lestari dengan menggunakan algoritma *K-Means Clustering*. Proses segmentasi

divalidasi melalui metode Elbow untuk menentukan jumlah kluster optimal dan Silhouette Score untuk mengevaluasi kualitas pemodelan. Hasil segmentasi kemudian divisualisasikan dalam bentuk grafik seperti scatter plot, pie chart, dan bar chart guna mendukung proses penargetan pembeli secara lebih efektif dan terarah. Temuan dari penelitian ini diharapkan dapat membantu tim pemasaran dalam memahami karakteristik setiap segmen pelanggan secara mendalam dan menyusun strategi pemasaran yang lebih tepat sasaran.

Kontribusi utama dari penelitian ini terletak pada penerapan gabungan metode clustering dan visualisasi multivariabel terhadap data historis pelanggan real estate selama satu dekade (2014–2024), khususnya dalam konteks wilayah berkembang seperti Kabupaten Jember, yang hingga saat ini masih jarang menjadi fokus penelitian serupa. Dengan pendekatan berbasis data ini, perusahaan diharapkan dapat meningkatkan efektivitas pengambilan keputusan strategis dan daya saing di industri properti.

II. METODE



Penelitian ini menggunakan pendekatan deskriptif kuantitatif dengan mengadopsi kerangka CRISP-DM (Cross-Industry Standard Process for Data Mining) dan menganalisis data sekunder dari 403 transaksi pelanggan PT Sembilan Bintang Lestari di perumahan Rembangan Hill Residence (2014–2024). Variabel yang digunakan meliputi usia, pendapatan, pekerjaan, tipe rumah, metode pembayaran, dan status pernikahan. Penelitian ini bertujuan untuk mengelompokkan pelanggan berdasarkan karakteristik demografi dan preferensi pembelian menggunakan K-Means Clustering, menentukan jumlah cluster optimal melalui metode Elbow, serta melakukan penargetan pembeli guna meningkatkan efektivitas penjualan properti.

2.1 Business Understanding

Tahap *Business Understanding* bertujuan untuk menyelaraskan tujuan analisis data dengan kebutuhan bisnis yang nyata, dalam hal ini membantu PT Sembilan Bintang Lestari meningkatkan efektivitas strategi pemasaran melalui pemahaman yang lebih dalam terhadap karakteristik pelanggan. Permasalahan

utama yang dihadapi perusahaan adalah belum adanya segmentasi pelanggan yang terstruktur, sehingga promosi yang dilakukan bersifat umum dan kurang tepat sasaran. Oleh karena itu, penelitian ini difokuskan untuk menyusun strategi pemasaran berbasis data dengan cara mengidentifikasi pola dan kecenderungan pelanggan melalui metode K-Means Clustering. Variabel yang digunakan meliputi usia, pendapatan, pekerjaan, metode pembayaran, tipe rumah, dan status pernikahan. Untuk menentukan jumlah kluster optimal, digunakan metode Elbow yang mengevaluasi nilai Within-Cluster Sum of Squares (WCSS) dan mengidentifikasi titik siku pada grafik sebagai indikator jumlah kluster terbaik. Hasil dari segmentasi ini diharapkan tidak hanya menghasilkan pembagian kelompok pelanggan yang homogen, tetapi juga memberikan dasar yang kuat bagi pengambilan keputusan strategis, seperti penyesuaian konten promosi, penawaran harga, maupun pendekatan komunikasi yang lebih personal. Dengan demikian, pemahaman bisnis yang diperoleh dari proses segmentasi ini akan menjadi landasan penting dalam merancang strategi pemasaran yang lebih adaptif dan berbasis data, sesuai dengan kebutuhan pasar properti yang semakin kompetitif..

2.3 Data Understanding

Data yang digunakan dalam penelitian ini diperoleh dari sistem internal PT Sembilan Bintang Lestari dan mencakup informasi transaksi pelanggan selama periode 2014 hingga 2024 pada proyek perumahan Rembangan Hill Residence. Sebelum dilakukan analisis lebih lanjut, data dieksplorasi secara menyeluruh untuk memahami struktur, distribusi, dan kualitasnya. Proses eksplorasi ini meliputi identifikasi jumlah entri, jenis atribut, konsistensi data, serta deteksi terhadap nilai hilang, duplikasi, dan potensi outlier. Dataset terdiri dari enam atribut utama yang dianggap relevan untuk proses segmentasi pelanggan, yaitu usia, pendapatan, pekerjaan, tipe rumah, metode pembayaran, dan status pernikahan. Atribut-atribut ini dipilih karena dinilai berkontribusi signifikan dalam merepresentasikan karakteristik demografis dan perilaku pembelian pelanggan. Usia dan pendapatan mencerminkan dimensi ekonomi dan tahap hidup pelanggan, sementara pekerjaan dan status pernikahan memberikan gambaran sosial dan preferensi hidup. Tipe rumah dan metode pembayaran berkaitan langsung dengan pilihan dan kemampuan dalam bertransaksi. Keseluruhan atribut ini akan digunakan untuk mengidentifikasi pola keterkaitan yang mendalam antar variabel guna menghasilkan segmentasi pelanggan yang bermakna. Pemahaman yang baik terhadap struktur data pada tahap awal ini sangat penting untuk memastikan bahwa proses clustering dengan algoritma K-Means dapat berjalan dengan optimal dan menghasilkan kluster yang akurat dan aplikatif bagi pengambilan keputusan pemasaran.

2.3 Data Preparation

Tahap Data Preparation bertujuan untuk menyiapkan data agar optimal digunakan dalam proses clustering. Langkah ini mencakup seleksi atribut, pembersihan data, transformasi, dan normalisasi agar data memiliki kualitas yang baik serta dapat diproses dengan algoritma K-Means yang berbasis jarak Euclidean.

Pertama, dilakukan seleksi terhadap fitur yang relevan. Dari data transaksi pelanggan perumahan Rembangan Hill Residence tahun 2014–2024, dipilih enam variabel utama: usia, pendapatan, pekerjaan, tipe rumah, metode pembayaran, dan status pernikahan. Kolom lain seperti nama, jenis kelamin, dan tanggal lahir dihapus karena tidak berkontribusi terhadap proses segmentasi.

Kedua, dilakukan pembersihan data dengan menghapus missing values dan data duplikat untuk menghindari noise. Selain itu, dilakukan deteksi dan penanganan outlier pada variabel numerik seperti usia dan pendapatan menggunakan metode Interquartile Range (IQR) dan capping agar nilai ekstrim tidak mendistorsi hasil pengelompokan.

Ketiga, transformasi data dilakukan terhadap variabel kategorikal. Dari enam variabel utama, empat di antaranya berskala kategorikal, yaitu pekerjaan, tipe rumah, metode pembayaran, dan status pernikahan. Keempat variabel ini diubah menjadi bentuk numerik menggunakan teknik Label Encoding karena algoritma K-Means hanya menerima input numerik. Misalnya, metode pembayaran "KPR" diberi label 1, sedangkan "Cash" diberi label 0.

Namun, penggunaan label encoding memiliki potensi bias numerik karena mengasumsikan adanya hubungan ordinal antar kategori, padahal nilai-nilai kategorikal bersifat nominal. Menurut [19], perhitungan jarak Euclidean tidak cocok digunakan untuk data kategorikal karena tidak adanya jarak alami antar kategori. Hal ini dapat menimbulkan distorsi dalam proses pengelompokan, di mana algoritma menganggap kategori tertentu lebih besar atau lebih kecil dari yang lain secara matematis.

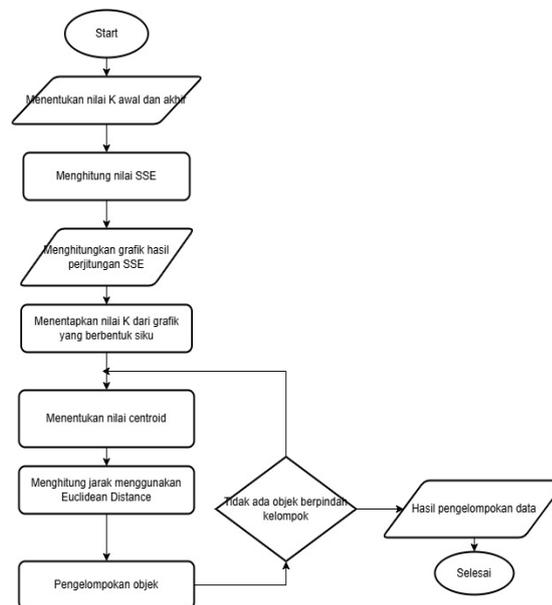
Untuk mengatasi hal tersebut, dilakukan validasi tambahan pada hasil clustering menggunakan Silhouette Score dan visualisasi multivariat (seperti boxplot, barchart, dan pie chart) untuk memastikan bahwa hasil segmentasi tetap relevan secara bisnis dan tidak terdistorsi oleh encoding semu.

Terakhir, seluruh variabel numerik, termasuk hasil encoding, dinormalisasi menggunakan Min-Max Scaling untuk mengubah rentang nilai menjadi antara 0 dan 1. Hal ini penting untuk mencegah dominasi variabel berskala besar, seperti pendapatan, terhadap variabel berskala kecil, seperti usia. Normalisasi terbukti penting dalam praktik data mining karena fitur dengan rentang nilai besar dapat secara tidak proporsional memengaruhi hasil analisis berbasis jarak

seperti K-Means [20], [21], [22]. Dengan skala yang seragam, semua variabel memiliki pengaruh yang seimbang dalam proses clustering.

2.4 Modelling

Penelitian ini menggunakan Algoritma Clustering K-means sebagai metode clustering. Untuk menentukan jumlah k yang paling ideal, gunakan pendekatan iteratif menggunakan metode Elbow [7]. Proses implementasi program menggunakan Algoritma K-Means Clustering digambarkan dalam flowchart berikut:



Gambar 1 Flowchart implementasi K-means dan Elbow method

Algoritma K-Means digunakan untuk mengelompokkan data ke dalam K kluster (dari K=2 hingga K=10) melalui serangkaian langkah iteratif. Proses dimulai dengan memilih K titik data secara acak sebagai centroid awal. Setiap titik data kemudian diukur jaraknya ke semua centroid menggunakan jarak Euclidean, yang dihitung dengan rumus:

$$D_{e} = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Berdasarkan jarak terdekat hasil perhitungan, data dialokasikan ke kluster yang sesuai dengan centroid tersebut. Setelah pengelompokan awal, nilai centroid diperbarui dengan menghitung rata-rata nilai data dalam kluster menggunakan persamaan:

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

Proses iterasi (menghitung jarak, mengelompokkan data, dan memperbarui centroid) diulang hingga tidak ada perubahan anggota kluster, menandakan bahwa algoritma telah mencapai konvergensi. Hasil akhirnya adalah pembagian data ke dalam K kluster yang

homogen, di mana setiap kluster memiliki centroid sebagai pusat representatifnya.

2.5 Evaluation

Penentuan jumlah kluster optimal dilakukan menggunakan metode Elbow dengan menghitung *Within-Cluster Sum of Squares (WCSS)* untuk rentang

$K=2$ hingga $K=10$ WCSS didefinisikan sebagai total jarak kuadrat antara setiap data (x_i) dengan centroid kluster tempatnya (c_j), dirumuskan sebagai berikut :

$$WCSS = \sum_{j=1}^k \sum_{x_i \in c_j} \text{jarak}(x_i, c_j)^2$$

di mana c_j merepresentasikan himpunan data dalam kluster ke- j . Nilai WCSS yang rendah menunjukkan kluster yang padat dan homogen. Hasil perhitungan WCSS divisualisasikan dalam grafik untuk mengidentifikasi titik optimal (elbow point), yaitu nilai K di mana penurunan WCSS mulai melambat signifikan. Selain itu, validasi kualitas kluster dilakukan dengan mengukur Silhouette Score, metrik yang dirumuskan sebagai:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

dengan $a(i)$ sebagai rata-rata jarak data ke- i ke anggota kluster sendiri, dan $b(i)$ sebagai rata-rata jarak data ke- i ke kluster terdekat lainnya. Skor ini berkisar antara -1 hingga 1 , dimana nilai mendekati 1 mengindikasikan struktur kluster yang terpisah jelas dan kohesif. Kombinasi analisis WCSS dan Silhouette Score memastikan pemilihan jumlah kluster yang optimal.

Validasi visual dilakukan melalui diagram batang (distribusi usia), boxplot (sebaran pendapatan), dan pie chart (preferensi pembayaran), yang menunjukkan konsistensi karakteristik tiap cluster. Interpretasi bisnis difokuskan pada penargetan pembeli dengan menganalisis profil unik setiap cluster.

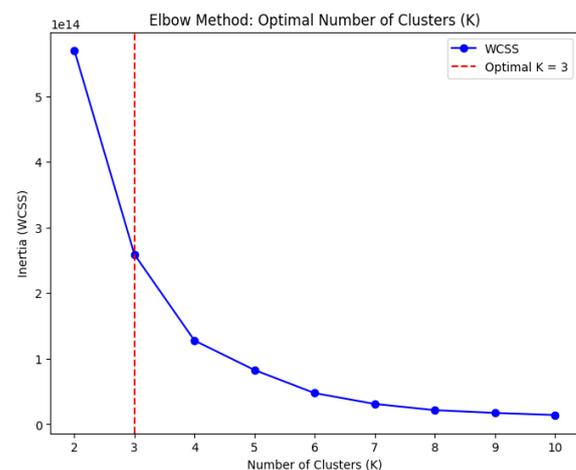
III. HASIL DAN PEMBAHASAN

3.1 Modelling

Proses clustering dalam penelitian ini dilakukan dengan menggunakan algoritma K-Means Clustering, yang digunakan untuk mengelompokkan pelanggan berdasarkan karakteristik tertentu seperti usia, pendapatan, pekerjaan, tipe rumah, metode pembayaran, dan status pernikahan. Untuk menentukan jumlah kluster yang paling tepat, dilakukan serangkaian percobaan dengan variasi jumlah kluster (K) mulai dari 2 hingga 10. Tujuan dari percobaan ini adalah untuk mengevaluasi struktur data dan mencari konfigurasi kluster yang paling menggambarkan segmentasi pelanggan secara akurat. Pemilihan jumlah kluster yang tepat sangat penting

agar hasil segmentasi tidak hanya valid secara statistik, tetapi juga dapat diterapkan dalam konteks bisnis untuk mendukung pengambilan keputusan yang efektif.

Langkah pertama dalam menentukan jumlah kluster optimal adalah dengan menggunakan metode Elbow, yang bekerja dengan menghitung nilai *Within-Cluster Sum of Squares (WCSS)* pada masing-masing nilai K . WCSS menunjukkan seberapa baik titik data dalam kluster dikelompokkan di sekitar pusatnya (centroid), di mana nilai yang lebih rendah menunjukkan pengelompokan yang lebih baik. Hasil WCSS untuk setiap nilai K kemudian divisualisasikan dalam bentuk grafik, dengan sumbu X menunjukkan jumlah kluster dan sumbu Y menunjukkan nilai WCSS. Grafik tersebut akan menunjukkan penurunan WCSS yang signifikan di awal, lalu melandai setelah titik tertentu. Titik belokan atau siku pada grafik—yang disebut *elbow point*—menunjukkan nilai K yang optimal karena menjadi titik kompromi antara peningkatan kualitas kluster dan kompleksitas model. Dalam penelitian ini, $K = 3$ dipilih sebagai jumlah kluster optimal karena merupakan titik siku yang paling jelas pada grafik Elbow, dan dipastikan kembali dengan analisis Silhouette Score untuk mengonfirmasi kualitas pemisahan kluster.



Gambar 2 Method Elbow: Titik cluster optimal

Berdasarkan hasil analisis menggunakan metode Elbow, seperti yang ditampilkan pada Gambar 1, jumlah kluster optimal yang diperoleh dalam proses segmentasi pelanggan adalah $K = 3$. Metode Elbow dipilih karena mampu mengidentifikasi titik optimal di mana penambahan jumlah kluster tidak lagi memberikan pengurangan signifikan terhadap nilai *Within-Cluster Sum of Squares (WCSS)*. Titik “siku” pada grafik Elbow menunjukkan bahwa pada nilai $K = 3$, terdapat keseimbangan antara jumlah kluster dan tingkat homogenitas di dalam kluster, sehingga diputuskan sebagai jumlah kluster yang paling representatif terhadap struktur data.

Untuk memastikan bahwa pemilihan jumlah kluster tersebut tidak hanya optimal secara visual, dilakukan validasi tambahan menggunakan Silhouette Score, yang mengevaluasi sejauh mana setiap objek cocok dengan klasternya sendiri dibandingkan dengan kluster lain. Silhouette Score dihitung berdasarkan selisih antara rata-rata jarak intrakluster dan jarak terdekat ke kluster lain, dengan nilai berkisar antara -1 hingga 1. Semakin tinggi nilai Silhouette, semakin baik kualitas pemisahan kluster yang dihasilkan. Hasil evaluasi menunjukkan bahwa untuk $K = 3$, diperoleh Silhouette Score sebesar 0,7643, sebagaimana tercantum pada Tabel 1. Nilai ini tergolong tinggi dan menunjukkan bahwa objek-objek dalam tiap kluster memiliki kesamaan yang kuat secara internal dan perbedaan yang jelas dibandingkan dengan kluster lainnya.

Dengan demikian, kombinasi antara metode Elbow dan validasi Silhouette Score memberikan dasar yang kuat bahwa pemilihan tiga kluster merupakan konfigurasi yang optimal untuk data pelanggan Rembulan Hill Residence. Kluster yang terbentuk dapat dikatakan memiliki struktur yang terpisah secara jelas dan representatif, yang pada akhirnya akan meningkatkan kualitas analisis segmentasi dan efektivitas dalam implementasi strategi pemasaran berbasis data.

Tabel 1 Hasil nilai Koefisien Silhouette

k	Silhouette Score
2	0.7888
3	0.7643
4	0.5592
5	0.6913
6	0.6914
7	0.7206
8	0.7121
9	0.6818
10	0.6978
11	0.6759

Berdasarkan hasil evaluasi, meskipun nilai Silhouette Score tertinggi diperoleh pada $K = 2$ (yaitu 0,7888), pemilihan $K = 3$ dianggap lebih tepat karena didukung oleh metode Elbow, yang menunjukkan titik siku pada $K = 3$ sebagai jumlah kluster optimal. Selain itu, nilai Silhouette Score untuk $K = 3$ sebesar 0,7643 masih tergolong tinggi, menandakan bahwa kluster yang terbentuk memiliki kualitas yang baik dan terpisah dengan jelas.

Pemilihan tiga kluster ini juga lebih relevan secara strategis, karena memungkinkan pembagian pelanggan ke dalam kelompok dengan karakteristik yang lebih beragam. Cluster 0, berisi 307 pelanggan, mewakili segmen pasar terbesar dengan karakteristik umum. Cluster 1 terdiri dari 72 pelanggan yang memiliki ciri

khas tersendiri, dan Cluster 2 dengan 24 pelanggan menunjukkan kelompok kecil namun potensial, seperti pembeli rumah premium.

Pembagian ini mempermudah perusahaan dalam menargetkan pembeli secara lebih tepat, merancang strategi pemasaran yang sesuai, serta meningkatkan efektivitas promosi. Dengan tiga segmen yang berbeda, strategi pemasaran dapat dibuat lebih personal dan sesuai dengan kebutuhan masing-masing kelompok pelanggan.

3.2. Evaluasi Model

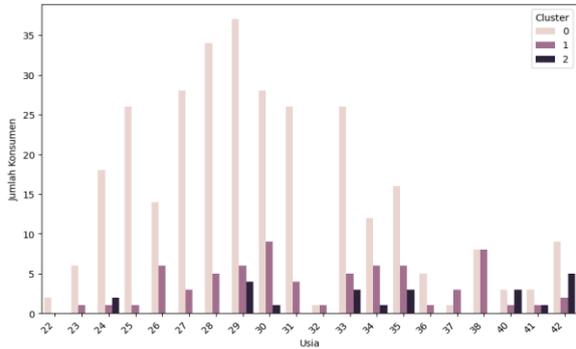
Evaluasi model dilakukan dengan membandingkan berbagai jumlah kluster berdasarkan dua metrik utama. Metode Elbow menunjukkan bahwa penurunan WCSS yang signifikan terjadi hingga $K = 3$, setelah itu penurunannya melambat, yang menandakan bahwa tiga kluster adalah jumlah optimal. Sementara itu, hasil evaluasi dengan Silhouette Score menunjukkan bahwa nilai tertinggi dicapai pada $K = 3$, dengan skor 0,7643. Ini menunjukkan bahwa hasil clustering memiliki pemisahan yang baik antara kluster satu dengan yang lainnya. Dengan kedua metode evaluasi ini, dipastikan bahwa jumlah kluster optimal adalah $K = 3$, karena memberikan keseimbangan antara kompleksitas model dan akurasi segmentasi pelanggan.

3.2. Visualisasi Hasil Clustering

Segmentasi pelanggan yang dilakukan dengan algoritma *K-Means* menghasilkan tiga cluster utama yang merepresentasikan kelompok pelanggan dengan karakteristik yang berbeda-beda. Untuk mempermudah pemahaman terhadap struktur dan perbedaan antar kluster, digunakan beberapa jenis visualisasi data seperti *boxplot*, diagram batang (*bar chart*), dan diagram lingkaran (*pie chart*). Visualisasi ini bertujuan untuk memberikan gambaran yang jelas mengenai distribusi dan kecenderungan masing-masing variabel penting dalam tiap kluster, seperti usia, pendapatan, pekerjaan, tipe rumah, metode pembayaran, dan status pernikahan.

Visualisasi ini tidak hanya memudahkan interpretasi hasil segmentasi, tetapi juga memberikan dukungan kuat bagi tim pemasaran dalam menyusun strategi yang tepat untuk setiap segmen pasar. Menurut [23], representasi grafis dari hasil klusterisasi dapat mempercepat proses pengambilan keputusan dalam bisnis, karena mampu menyampaikan informasi yang kompleks secara lebih sederhana, visual, dan komunikatif.

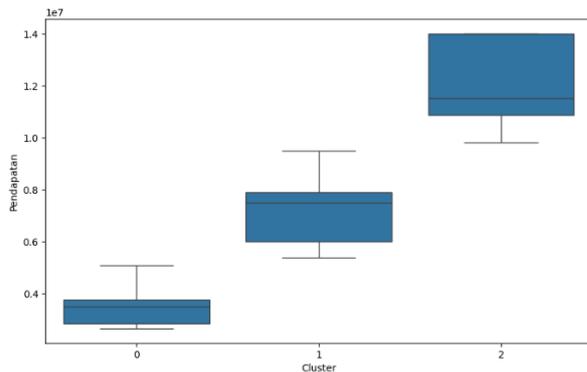
• Distribusi Usia Setiap Cluster



Gambar 3 Tampilan Barchart distribusi usia setiap cluster

Grafik distribusi usia konsumen menunjukkan bahwa **Cluster 0** memiliki jumlah konsumen tertinggi, terutama pada rentang usia produktif 28 hingga 30 tahun, dengan puncaknya di usia 29 tahun. **Cluster 1** memiliki konsentrasi yang lebih besar pada usia 30 hingga 38 tahun, yang mencerminkan konsumen pada tahap usia yang lebih matang. Sementara itu, **Cluster 2** memiliki jumlah konsumen paling sedikit, dengan sedikit peningkatan pada usia 41 hingga 42 tahun. Informasi ini membantu perusahaan memahami karakteristik usia di setiap cluster, sehingga dapat merancang strategi pemasaran yang lebih efektif dan terarah,

• Distribusi Pendapatan Setiap Cluster

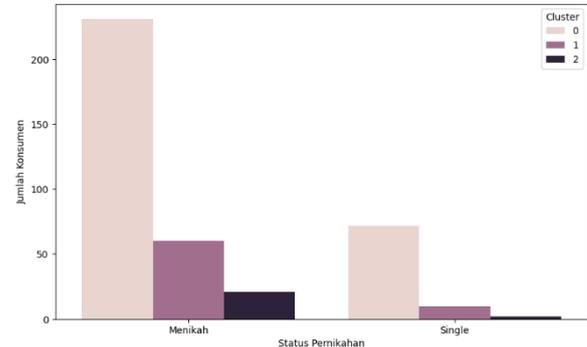


Gambar 4 Tampilan boxplot distribusi pendapatan setiap cluster

Grafik boxplot menunjukkan distribusi pendapatan di setiap cluster. Cluster 0 memiliki rentang pendapatan paling rendah, dengan median sekitar 4 juta rupiah, yang mengindikasikan kelompok dengan daya beli stabil. Cluster 1 menunjukkan pendapatan yang lebih tinggi, dengan median sekitar 8 juta rupiah, menandakan daya beli yang lebih baik dibandingkan Cluster 0. Sementara itu, Cluster 2 memiliki pendapatan tertinggi, dengan median sekitar 14 juta rupiah dan rentang pendapatan yang lebih luas, mencerminkan segmen pelanggan dengan kemampuan finansial yang jauh lebih besar. Analisis ini memberikan wawasan

penting untuk menyusun strategi pemasaran berdasarkan daya beli masing-masing cluster.

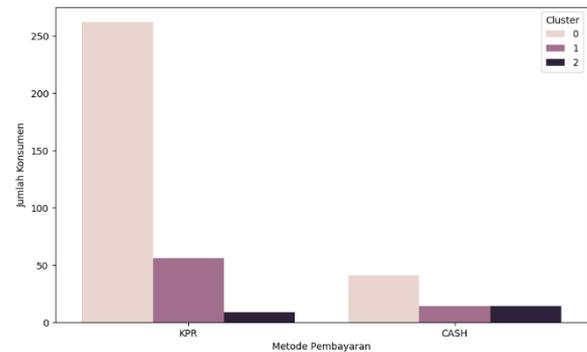
• Distribusi Status Pernikahan Setiap Cluster



Gambar 5 Tampilan Barchart Distribusi Pernikahan setiap cluster

Berdasarkan grafik, distribusi status pernikahan konsumen dapat dirangkum sebagai berikut: Di Cluster 0, terdapat lebih dari 200 konsumen yang menikah dan sekitar 50 konsumen single. Cluster 1 memiliki sekitar 60 konsumen yang menikah dan sekitar 10 konsumen single. Sementara itu, Cluster 2 mencatatkan sekitar 20 konsumen yang menikah dan kurang dari 10 konsumen single. Data ini menunjukkan dominasi konsumen menikah di semua cluster, dengan proporsi terbesar di Cluster 0.

• Distribusi Pembayaran Setiap Cluster



Gambar 6 Tampilan Barchart preferensi pembayaran setiap cluster

Grafik ini memperlihatkan distribusi metode pembayaran konsumen menurut cluster. Terlihat bahwa pada cluster 0 (merah muda), mayoritas konsumen menggunakan metode KPR dengan jumlah lebih dari 250 konsumen, sedangkan sekitar 50 konsumen memilih metode CASH. Pada cluster 1 (ungu), terdapat sekitar 50 konsumen yang menggunakan KPR dan 20 yang memilih CASH. Cluster 2 (ungu tua) menunjukkan jumlah konsumen yang lebih sedikit baik untuk metode KPR maupun CASH, namun konsumen metode CASH lebih banyak dibandingkan KPR. Informasi ini penting untuk memahami preferensi

Rembangsan Hill Residence dengan menganalisis 403 data pelanggan. Analisis dilakukan melalui tahapan CRISP-DM yang mencakup Business Understanding, Data Understanding, Data Preparation, Modeling, dan Evaluation. Dengan menggunakan metode Elbow untuk menentukan jumlah kluster optimal, diperoleh hasil bahwa jumlah kluster terbaik adalah tiga, yaitu Cluster 0 sebanyak 307 pelanggan, Cluster 1 sebanyak 72 pelanggan, dan Cluster 2 sebanyak 24 pelanggan. Segmentasi ini menunjukkan bahwa pelanggan dapat dikelompokkan secara efektif berdasarkan karakteristik yang relevan dan berbeda satu sama lain secara signifikan.

2. Clustering dengan algoritma K-Means efektif untuk mengelompokkan pelanggan berdasarkan variabel seperti usia, pendapatan, pekerjaan, tipe rumah, dan metode pembayaran. Teknik ini membantu perusahaan real estate memahami kebutuhan setiap segmen pelanggan sehingga tim pemasaran bisa menargetkan pembeli. Misalnya, pelanggan dengan pendapatan Rp2,7 juta hingga Rp5 juta di Cluster 0 lebih memilih rumah tipe 30 dengan skema KPR, sementara pelanggan berpendapatan di atas Rp9,8 juta di Cluster 2 lebih memilih rumah premium tipe 55 dan 70 dengan pembayaran tunai. Selain segmentasi, clustering juga berguna untuk menentukan lokasi pengembangan properti, merancang fasilitas yang sesuai, dan mengoptimalkan penawaran harga serta pembayaran. Visualisasi hasil clustering mempermudah analisis pola pelanggan, meningkatkan efisiensi pemasaran, dan memperkuat daya saing perusahaan di pasar yang kompetitif. Pendekatan ini membantu perusahaan merumuskan strategi yang lebih relevan dan berdampak positif pada kinerja bisnis.

4.2. Saran

Mengintegrasikan hasil segmentasi pelanggan ke dalam model Business Intelligence untuk pemahaman holistik dan wawasan perilaku pembelian.

DAFTAR PUSTAKA

- [1] "bank indonesia," 2024.
- [2] A. Sudarso, *Konsep E-Bisnis*, vol. 01, no. November, 2020.
- [3] Y. Yang, "Application of Big Data in Real Estate Marketing," 2019.
- [4] "Company profile PT. Sembilan Bintang Lestari," 2022.
- [5] A. Burhan, H. Kiat, Y. Azhar, and V. Rahmayanti, "Penerapan Metode K-Means Dengan Metode Elbow Untuk Segmentasi Pelanggan Menggunakan Model RFM (Recency, Frequency & Monetary)," *REPOSITOR*, vol. 2, no. 7, pp. 945–952, 2020.
- [6] F. Tjiptono and G. Chandra, *Pemasaran Strategik*. 2013.
- [7] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *J. TEKNO KOMPAK*, vol. 15, no. 2, pp. 25–36, 2021.
- [8] J. Wijaya, H. P. Ipung, and M. A. Soetomo, "DESIGN AN AGILE OF MACHINE LEARNING TO PREDICTIVE HOUSE PRICING AND TARGETING SEGMENTED MARKET," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Sep. 2022. doi: 10.1145/3557738.3557856.
- [9] M. S. Cruising, "Data mining in customer segmentation," no. September, pp. 189–192, 2016.
- [10] C. Schröder, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 526–534. doi: 10.1016/j.procs.2021.01.199.
- [11] F. Barrera, M. Segura, and C. Maroto, "Multiple criteria decision support system for customer segmentation using a sorting outranking method," *Expert Syst. Appl.*, vol. 238, Mar. 2024, doi: 10.1016/j.eswa.2023.122310.
- [12] P. Y. Jadhav and D. D. Parasar, "Customer Segmentation and Buyer Targeting Approach," *Int. J. Recent Technol. Eng.*, vol. 9, no. 4, pp. 295–303, Nov. 2020, doi: 10.35940/ijrte.D5013.119420.
- [13] Erlin Windia Ambarsari, Dedin Fathudin, and Gravita Alfiani, "Utilizing K-Means Clustering to Understanding Audience Interest in SEO-Optimized Media Content," *J. Comput. Informatics Res.*, vol. 3, no. 2, pp. 208–214, 2024, doi: 10.47065/comforch.v3i2.1207.
- [14] I. Herdiana, M. A. Kamal, Triyani, M. N. Estri, and Renny, "A More Precise Elbow Method for Optimum K-means Clustering," pp. 1–22, 2025, [Online]. Available: <http://arxiv.org/abs/2502.00851>
- [15] R. Huang and S. Mao, "Research on Precision Marketing of Real Estate Market Based on Data Mining," *Sci. Program.*, vol. 2022, 2022, doi: 10.1155/2022/8198568.

- [16] E. A. E. Dawood, E. Elfakhrany, and F. A. Maghraby, "Improve profiling bank customer's behavior using machine learning," *IEEE Access*, vol. 7, pp. 109320–109327, 2019, doi: 10.1109/ACCESS.2019.2934644.
- [17] R. Ananda and A. Z. Yamani, "Terakreditasi SINTA Peringkat 2 Penentuan Centroid Awal K-means pada proses Clustering Data Evaluasi Pengajaran Dosen," *masa berlaku mulai*, vol. 1, no. 3, pp. 544–550, 2020.
- [18] A. Al Fahrenzi, F. Insani, E. Budianita, and I. Afrianty, "Implementasi Algoritma K-Means dalam Menentukan Clustering pada Penilaian Kepuasan Pelanggan di Badan Pelatihan Kesehatan Pekanbaru," *Indones. J. Innov. Multidisipliner Res.*, vol. 1, no. 4, pp. 474–492, 2023, doi: 10.31004/ijim.v1i4.53.
- [19] V. Kumar, S. Boriah, and V. Chandola, "Similarity measures for categorical data: A comparative evaluation," *Soc. Ind. Appl. Math. - 8th SIAM Int. Conf. Data Min. 2008, Proc. Appl. Math. 130*, vol. 1, no. June, pp. 243–254, 2008, doi: 10.1137/1.9781611972788.22.
- [20] P. Dalatu, A. Fitrianto, and A. Mustapha, "Improved normalization and standardization techniques for higher purity in K-means clustering," *Far East J. Math. Sci.*, vol. 100, pp. 859–871, 2016, doi: 10.17654/MS100060859.
- [21] M. D. V. Prasad and S. T, "Clustering Accuracy Improvement Using Modified Min-Max Normalization Technique Min-Max Normalization Technique," pp. 0–7, 2024, doi: 10.20944/preprints202411.0486.v1.
- [22] Y. Dimas Pratama and A. Salam, "Comparison of Data Normalization Techniques on KNN Classification Performance for Pima Indians Diabetes Dataset," *J. Appl. Informatics Comput.*, vol. 9, no. 3, p. 693, 2025, [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [23] Y. Yuan, H. Xu, M. Krishnamurthy, and P. Vijayakumar, "Visualization analysis of educational data statistics based on big data mining," *J. Comput. Methods Sci. Eng.*, vol. 24, no. 3, pp. 1785–1793, 2024, doi: 10.3233/JCM-230003.