

EKSPLORASI DATA PENJUALAN DI PT XYZ UNTUK MENGUNGKAP FAKTA-FAKTA SEGMENTASI KONSUMEN PROPERTI MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING

Stefanus¹, Darmansah²

¹Mahasiswa Program Studi Sistem Informasi, Universitas Putera Batam

²Dosen Program Studi Sistem Informasi, Universitas Putera Batam

email: pb221510022@upbatam.ac.id

ABSTRACT

Housing sales data in property companies are commonly utilized only for administrative purposes, limiting their potential to generate strategic insights into consumer behavior and purchasing patterns. PT XYZ possesses a substantial volume of housing sales transaction data, yet this data has not been systematically analyzed to support data-driven marketing decisions. This study aims to identify consumer segmentation in the property sector at PT XYZ by applying the K-Means Clustering algorithm. A quantitative descriptive research approach was employed using housing sales data from January 2022 to June 2025. The variables analyzed include housing selling price, housing unit type, and payment method. Data processing followed the Knowledge Discovery in Databases (KDD) framework, while the optimal number of clusters was determined using the Elbow Method and Silhouette Coefficient. The Clustering process was conducted using RapidMiner and Google Colaboratory. The results reveal three distinct consumer clusters with different purchasing characteristics. These findings indicate that K-Means Clustering is effective for property consumer segmentation and provides meaningful insights to support more targeted and effective marketing strategies.

Keywords: *Consumer Segmentation, Data Mining, K-Means Clustering, Property.*

PENDAHULUAN

Industri properti perumahan memiliki peran strategis dalam perekonomian, baik sebagai pemenuhan kebutuhan dasar manusia maupun sebagai instrumen investasi jangka panjang. Seiring dengan meningkatnya aktivitas pembangunan dan penjualan perumahan, perusahaan properti menghasilkan data transaksi dalam jumlah besar yang mencerminkan perilaku dan preferensi konsumen. Namun, pada praktiknya, data penjualan perumahan masih sering dimanfaatkan

sebatas kebutuhan administratif dan pelaporan internal, tanpa dilakukan pengolahan lebih lanjut untuk menghasilkan informasi strategis.

PT XYZ merupakan perusahaan pengembang properti yang memiliki data transaksi penjualan rumah yang cukup besar dan bervariasi, meliputi informasi harga jual, tipe unit, serta metode pembayaran yang digunakan oleh konsumen. Sayangnya, data tersebut belum dianalisis secara sistematis untuk mengidentifikasi pola pembelian dan segmentasi konsumen. Akibatnya,

perusahaan belum memiliki gambaran yang jelas mengenai karakteristik kelompok pembeli yang berbeda, sehingga strategi pemasaran yang diterapkan cenderung bersifat umum dan kurang tepat sasaran.

Pemanfaatan teknik *Data Mining*, khususnya metode *Clustering*, dapat menjadi solusi untuk mengatasi permasalahan tersebut. *Clustering* memungkinkan pengelompokan konsumen berdasarkan kemiripan karakteristik pembelian tanpa memerlukan label kelas tertentu. Algoritma K-Means merupakan salah satu metode *Clustering* yang paling banyak digunakan karena kesederhanaan, efisiensi, serta kemampuannya dalam mengolah data berukuran besar. (Maulani et al., 2025)

Beberapa penelitian terdahulu seperti penelitian (Suderajat et al., 2025) menunjukkan bahwa K-Means *Clustering* efektif dalam membentuk segmentasi pasar properti berdasarkan harga, luas bangunan, maupun preferensi konsumen. Namun, penerapan algoritma ini pada data penjualan perumahan lokal dengan menggabungkan variabel harga jual, tipe unit, dan metode pembayaran masih relatif terbatas. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma K-Means *Clustering* dalam mengelompokkan konsumen properti di PT XYZ, sehingga diperoleh segmentasi konsumen yang dapat mendukung pengambilan keputusan pemasaran secara lebih tepat dan berbasis data.

KAJIAN TEORI

2.1. Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases (KDD) adalah proses yang sistematis untuk mengekstraksi pengetahuan yang

bernilai dari kumpulan data besar. KDD melibatkan beberapa tahapan, mulai dari pemilihan data, pembersihan, transformasi, *Data Mining*, hingga evaluasi hasil. KDD sangat penting dalam mengelola data besar yang dihasilkan dari berbagai aktivitas manusia, seperti transaksi digital dan interaksi pengguna. Berdasarkan penelitian yang dilakukan oleh (Ardhi Baskara et al., 2025), KDD menyediakan kerangka kerja yang memungkinkan organisasi untuk mengelola data yang kompleks, memanfaatkan berbagai teknik *Data Mining* untuk menemukan pola yang valid dan dapat diinterpretasikan, serta mendukung pengambilan keputusan berbasis data yang lebih baik. Dalam penelitian ini, KDD diterapkan untuk menganalisis data transaksi penjualan perumahan untuk menemukan pola pembelian konsumen yang tersembunyi yang dapat digunakan dalam strategi pemasaran.

2.2. Data Mining

Data Mining adalah proses eksplorasi dan analisis data untuk menemukan pola atau informasi tersembunyi yang dapat digunakan untuk mendukung pengambilan keputusan. Dalam konteks ini, *Data Mining* menggabungkan teknik statistik, pembelajaran mesin, dan kecerdasan buatan untuk mengolah data dalam jumlah besar dan kompleks. Algoritma K-Means *Clustering*, yang merupakan salah satu metode dalam *Data Mining*, digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik tanpa memerlukan informasi label terlebih dahulu. Berdasarkan penelitian oleh (Darmansah et al., 2024), K-Means *Clustering* efektif digunakan dalam berbagai sektor industri, termasuk manufaktur, untuk segmentasi tenaga

kerja. Dengan menggunakan data variabel seperti usia, pendidikan, pengalaman kerja, gaji, dan keterampilan, perusahaan dapat mengidentifikasi segmen-segmen pekerja yang memiliki karakteristik serupa, yang memungkinkan pengelolaan sumber daya manusia yang lebih efisien dan berbasis data.

2.3. Perumahan dan Properti

Menurut Undang-Undang No. 1 Tahun 2011, Perumahan adalah kumpulan rumah yang dilengkapi dengan sarana, prasarana, dan utilitas umum, yang berfungsi tidak hanya sebagai tempat tinggal tetapi juga menunjang aktivitas kehidupan masyarakat baik di kawasan perkotaan maupun perdesaan. (Alinda et al., 2021)

2.4. Clustering

Clustering adalah teknik dalam *Data Mining* yang digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristik antar data. Dalam metode *unsupervised learning*, algoritma *K-Means Clustering* membagi data ke dalam *K* cluster, dengan tujuan memaksimalkan kesamaan dalam satu cluster dan meminimalkan kesamaan antar cluster. Teknik ini banyak digunakan dalam segmentasi pasar, seperti pada penelitian ini yang mengelompokkan tenaga kerja berdasarkan usia, pendidikan, dan keterampilan. (Darmansah & Wardani W.N, 2022)

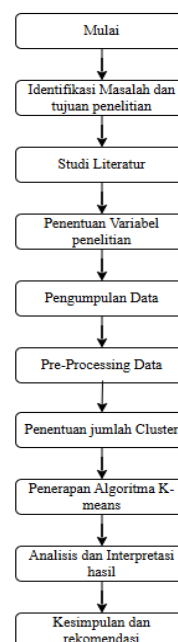
2.5. Algoritma K-Means

Algoritma *K-Means* adalah metode *Clustering* yang membagi data menjadi beberapa kelompok berdasarkan kesamaan karakteristik. Proses ini dimulai dengan memilih jumlah cluster (*K*) dan mengalokasikan data ke kelompok berdasarkan kedekatannya dengan pusat *cluster* (*centroid*). Algoritma ini bekerja secara iteratif

hingga posisi *centroid* tidak berubah. *K-Means* sering digunakan untuk mengelompokkan data besar secara efisien, seperti pada penelitian ini yang mengelompokkan penyebaran COVID-19 ke dalam kategori rendah, sedang, dan tinggi. (Darmansah, 2021)

2.6. Metode *Elbow*

Metode *Elbow* digunakan untuk menentukan jumlah cluster optimal dengan menganalisis perubahan nilai *Sum of Squared Errors* (SSE) terhadap jumlah cluster yang dibentuk. SSE menunjukkan total jarak kuadrat antara setiap data dan pusat cluster terdekatnya. Jumlah cluster optimal ditentukan pada titik "siku" (*elbow point*), yaitu saat penurunan nilai SSE mulai melambat, yang menandakan bahwa penambahan cluster selanjutnya tidak lagi meningkatkan kualitas cluster secara signifikan (Permadi et al., 2023).



Penggunaan metode *Elbow* membantu mengurangi subjektivitas dalam penentuan jumlah kluster pada algoritma K-Means. (Qusyairi et al., 2024) menyatakan bahwa metode *Elbow* efektif dalam menyeimbangkan kompleksitas model dan kualitas klusterisasi, sehingga banyak digunakan sebagai tahap awal sebelum proses evaluasi kluster lebih lanjut dilakukan.

2.7. Metode Silhouette Coefficient

Metode Silhouette Coefficient digunakan untuk mengevaluasi kualitas kluster dengan mengukur tingkat kesamaan data dalam satu kluster dibandingkan dengan kluster lainnya. Nilai Silhouette berada pada rentang -1 hingga 1 , di mana nilai yang mendekati 1 menunjukkan bahwa data memiliki tingkat kohesi dan separasi kluster yang baik. Metode ini memberikan ukuran kuantitatif terhadap kualitas hasil klusterisasi yang terbentuk (Febby Arisca Zurfani et al., 2024).

Silhouette Coefficient sering digunakan sebagai metode validasi untuk memperkuat hasil penentuan jumlah kluster menggunakan metode *Elbow*. Penelitian oleh (Lashiyanti et al., 2023) menunjukkan bahwa kombinasi kedua metode tersebut mampu menghasilkan pemilihan jumlah kluster yang lebih akurat dan memiliki interpretasi yang lebih jelas secara praktis.

METODE PENELITIAN

3.1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode deskriptif. Fokus penelitian diarahkan pada penerapan algoritma K-Means *Clustering* untuk mengelompokkan pembeli rumah berdasarkan tiga variabel utama, yaitu harga jual, tipe unit yang dipilih dan metode pembayaran yang digunakan. Pendekatan kuantitatif dipilih karena

penelitian melibatkan pengolahan data numerik untuk menghasilkan pola yang objektif dan dapat diinterpretasikan.

Gambar 3.1 Desain Penelitian

3.2. Variabel Penelitian

Variabel yang dianalisis dalam penelitian ini terdiri dari:

1. Harga Jual Rumah: Data harga rumah yang mencakup rumah dengan harga terjangkau hingga premium.
2. Tipe Unit Rumah: Variabel ini mencakup tipe unit rumah seperti tipe 36, tipe 45, dan tipe 100 yang mewakili ukuran rumah yang dibeli.
3. Metode Pembayaran: Metode pembayaran yang digunakan oleh konsumen, baik itu Cash Keras, Cash Bertahap, atau Kredit Pemilikan Rumah (KPR).

3.3. Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh data transaksi penjualan rumah yang tercatat di PT XYZ, dengan periode data dari Januari 2022 hingga Juni 2025. Sampel yang digunakan adalah transaksi rumah dengan status aktif, belum lunas, dan sudah lunas, dengan mengabaikan transaksi yang dibatalkan. Data yang digunakan terdiri dari 890 baris data setelah seleksi.

3.4. Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari PT XYZ. Data tersebut meliputi informasi transaksi penjualan rumah, termasuk harga jual rumah, tipe unit rumah, dan metode pembayaran yang digunakan oleh konsumen. Pengumpulan data dilakukan melalui database perusahaan dan rekaman transaksi penjualan yang telah diproses dan disimpan dengan baik.

3.5. Metode Analisis Data

Proses analisis data mengikuti tahapan KDD, yang dimulai dengan seleksi data untuk memilih data yang relevan, dilanjutkan dengan pembersihan data untuk menghapus duplikasi atau data yang tidak valid, dan transformasi data dengan mengubah variabel kategorikal menjadi numerik agar dapat diproses oleh algoritma. Setelah itu, algoritma K-Means *Clustering* diterapkan menggunakan *RapidMiner* untuk mengelompokkan konsumen ke dalam kluster berdasarkan kesamaan karakteristik pembelian.

Metode *Elbow* digunakan untuk menentukan jumlah kluster optimal dengan melihat titik "*elbow*" pada grafik penurunan *Sum of Squared Errors* (SSE), sedangkan *Silhouette Coefficient* digunakan untuk mengukur kualitas pemisahan kluster.

HASIL DAN PEMBAHASAN

4.1. Hasil Pengolahan Data

Data penelitian diperoleh dari transaksi penjualan rumah di PT XYZ selama periode Januari 2022 hingga Juni 2025. Data awal berjumlah 3.643 transaksi yang mencakup informasi harga jual rumah, tipe unit, metode pembayaran, dan status transaksi. Selanjutnya dilakukan proses seleksi data dengan mengacu pada batasan penelitian, yaitu hanya menggunakan transaksi berstatus aktif, belum lunas, dan lunas, serta mengecualikan transaksi dengan status batal. Hasil seleksi tersebut menghasilkan 890 data transaksi yang digunakan sebagai dataset penelitian.

Pada tahap pembersihan data, dilakukan pemeriksaan terhadap kelengkapan dan konsistensi data. Seluruh data yang digunakan telah memenuhi variabel penelitian, sehingga tidak ditemukan data kosong maupun duplikasi yang dapat memengaruhi proses analisis. Dengan

demikian, data dinyatakan layak untuk digunakan pada tahap selanjutnya.

Tahap transformasi data dilakukan untuk menyesuaikan format variabel dengan kebutuhan algoritma K-Means. Variabel harga jual rumah sudah berbentuk numerik, sedangkan variabel tipe unit rumah dan metode pembayaran yang bersifat kategorikal diubah ke dalam bentuk numerik. Transformasi ini bertujuan agar seluruh variabel dapat dihitung menggunakan jarak *Euclidean* pada proses klusterisasi.

Untuk menghindari perbedaan skala antarvariabel, seluruh data kemudian dinormalisasi menggunakan metode *Min-Max Normalization* sehingga setiap atribut berada pada rentang nilai 0 hingga 1. Proses normalisasi ini bertujuan untuk memastikan bahwa setiap variabel memiliki kontribusi yang seimbang dalam pembentukan kluster. Dataset hasil pengolahan selanjutnya digunakan pada tahap penentuan jumlah kluster optimal dan penerapan algoritma K-Means *Clustering*.

4.2. Penentuan Jumlah Kluster Optimal

Metode *Elbow* dilakukan dengan menghitung nilai *Sum of Squared Errors* (SSE) pada beberapa nilai K, yaitu K = 2 hingga K = 6. SSE merupakan jumlah kuadrat jarak antara setiap data dengan pusat kluster (*centroid*) terdekat. Semakin kecil nilai SSE, maka semakin homogen kluster yang terbentuk.

Hasil perhitungan menunjukkan bahwa nilai SSE mengalami penurunan yang signifikan dari K = 2 ke K = 3. Namun, setelah K = 3, penurunan nilai SSE cenderung melandai dan tidak menunjukkan perbaikan yang berarti. Pola tersebut membentuk titik siku (*elbow point*) pada K = 3, yang mengindikasikan bahwa penambahan jumlah kluster di atas nilai tersebut tidak lagi memberikan

peningkatan kualitas kluster secara signifikan.

Berdasarkan hasil tersebut, jumlah kluster optimal yang direkomendasikan oleh Metode *Elbow* adalah tiga kluster.

Untuk memperkuat hasil Metode *Elbow*, dilakukan evaluasi menggunakan *Silhouette Coefficient*. Metode ini mengukur tingkat kesamaan data dalam satu kluster dibandingkan dengan kluster lain, dengan nilai berada pada rentang -1 hingga 1. Nilai *Silhouette* yang mendekati 1 menunjukkan bahwa data berada pada kluster yang tepat dan memiliki pemisahan yang baik.

Hasil pengujian *Silhouette Coefficient* menunjukkan bahwa nilai rata-rata *Silhouette* tertinggi diperoleh pada $K = 3$, sehingga mendukung hasil yang diperoleh dari Metode *Elbow*. Dengan demikian, pembentukan tiga kluster dinilai memiliki tingkat homogenitas internal dan separasi antar kluster yang paling optimal.

4.3. Interpretasi Hasil Klusterisasi

1. Kluster 1: Pembeli rumah dengan harga menengah dan tipe unit

menengah, dengan preferensi pembayaran cash bertahap. Kelompok ini mencerminkan konsumen yang menginginkan fleksibilitas pembayaran namun tetap dalam batasan harga menengah.

2. Kluster 2: Pembeli rumah premium dengan tipe besar dan pilihan pembayaran cicilan uang muka. Kelompok ini terdiri dari konsumen yang lebih mampu secara finansial dan cenderung mencari rumah dengan fasilitas lebih besar.

3. Kluster 3: Pembeli rumah dengan harga lebih rendah dan tipe unit kecil, lebih banyak memilih KPR. Kelompok ini menggambarkan konsumen yang lebih terbatas dalam hal anggaran dan mengandalkan fasilitas pembiayaan bank.

Tabel 4.1 Analisa hasil Klusterisasi

No.	Kluster	Jumlah data	Rata-rata (cara pembayaran)	Rata-rata (Harga jual)	Rata-rata (Tipe unit)
1	Kluster 1	186	1,839	0,209	0,253
2	Kluster 2	68	3,000	0,218	0,257
3	Kluster 3	636	4,000	0,036	0,071

SIMPULAN

Berdasarkan hasil analisis data penjualan rumah di PT XYZ menggunakan algoritma *K-Means Clustering* terhadap 890 data transaksi, penelitian ini berhasil membentuk tiga kluster konsumen dengan karakteristik pembelian yang berbeda. Penentuan jumlah kluster optimal

dilakukan menggunakan Metode *Elbow*, yang menunjukkan titik siku pada $K = 3$, dan diperkuat dengan evaluasi *Silhouette Coefficient* yang memberikan nilai terbaik pada jumlah kluster yang sama. Hasil ini menunjukkan bahwa pembentukan tiga kluster memiliki tingkat pemisahan dan homogenitas data yang optimal.

Hasil klusterisasi menunjukkan bahwa Klaster 1 terdiri dari 186 data transaksi (20,90%), yang didominasi oleh pembeli rumah dengan harga dan tipe unit kategori menengah serta kecenderungan menggunakan metode pembayaran cash bertahap. Klaster 2 merupakan klaster dengan jumlah anggota paling sedikit, yaitu 68 data transaksi (7,64%), yang merepresentasikan pembeli rumah dengan harga dan tipe unit lebih tinggi dengan preferensi metode pembayaran cicilan uang muka. Sementara itu, Klaster 3 menjadi klaster terbesar dengan 636 data transaksi (71,46%), yang didominasi oleh pembeli rumah dengan harga dan tipe unit lebih rendah serta metode pembayaran Kredit Pemilikan Rumah (KPR).

Hasil tersebut menunjukkan bahwa algoritma K-Means *Clustering* mampu mengelompokkan konsumen properti secara objektif berdasarkan variabel harga jual, tipe unit, dan metode pembayaran. Segmentasi yang dihasilkan memberikan gambaran kuantitatif mengenai distribusi konsumen pada masing-masing segmen, sehingga dapat dimanfaatkan oleh perusahaan sebagai dasar dalam menyusun strategi pemasaran yang lebih terarah, khususnya dalam penentuan skema pembayaran, segmentasi produk, dan target pasar sesuai dengan karakteristik setiap klaster.

DAFTAR PUSTAKA

- Alinda, S. N., Setiawan, A. Y., & Sudrajat, A. (2021). Alih Fungsi Lahan Dari Sawah Menjadi Perumahan Di Kampung Gumuruh Desa Nagrak Kecamatan Canguang Kabupaten Bandung. *Geoarea*, 04(02), 55–67.
- Ardhi Baskara, A., Maharani Piranti, N., & Fahrury Romdendine, M. (2025). Framework Data Mining: Sebuah Survei. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4886–4895. <https://doi.org/10.36040/jati.v9i3.13803>
- Darmansah, D. D. (2021). Analisis Penyebaran Penularan Virus Covid-19 di Provinsi Jawa Barat Menggunakan Algoritma K-Means Clustering. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(3), 1188–1199. <https://doi.org/10.35957/jatisi.v8i3.1034>
- Darmansah, Fitra Kasma Putra, & Tomy Nanda Putra. (2024). Penerapan Algoritma K-Means Clustering Pada Segmentasi Tenaga Kerja di Batam Dalam Sektor Manufaktur. *JSAI (Journal Scientific and Applied Informatics)*, 7(3), 522–530. <https://doi.org/10.36085/jsai.v7i3.7209>
- Darmansah, & Wardani W.N. (2022). Analisis Penyebab Kerusakan Tanaman Cabai Menggunakan Metode K-Means. *Jurnal Teknik Informatika Dan Sistem Informasi*, 7(2), 126–134.
- Febby Arisca Zurfani, Sawaluddin, Mardinarsih, & Muhammad Romi Syahputra. (2024). Analisis Metode Clustering K-Means pada Zonasi Daerah Terdampak Banjir di Kota Medan dengan Evaluasi Silhouette Coefficient. *Algoritma : Jurnal Matematika, Ilmu Pengetahuan Alam, Kebumihan Dan Angkasa*, 2(6), 176–181. <https://doi.org/10.62383/algoritma.v2i6.270>
- Lashiyanti, A. R., Munthe, I. R., & ... (2023). Optimisasi Klusterisasi Nilai Ujian Nasional dengan Pendekatan

- Algoritma K-Means, Elbow, dan Silhouette. *Jurnal Ilmu Komputer* ..., 6, 14–20.
<http://ejournal.sisfokomtek.org/index.php/jikom/article/view/1550>
<http://ejournal.sisfokomtek.org/index.php/jikom/article/download/1550/1026>
- Maulani, V. R., Barata, M. A., & Yuwita, P. E. (2025). A Improving House Price Clustering Results with K-means through the Implementation of One-hot Encoding Pre-processing Technique. *Journal of Applied Informatics and Computing*, 9(3), 741–748.
<https://doi.org/10.30871/jaic.v9i3.9481>
- Permadi, V. A., Tahalea, S. P., & Agusdin, R. P. (2023). K-Means and Elbow Method for Cluster Analysis of Elementary School Data. *Progres Pendidikan*, 4(1), 50–57.
<https://doi.org/10.29303/prospek.v4i1.328>
- Qusyairi, M., Zul Hidayatullah, & Arnila Sandi. (2024). Penerapan K-Means Clustering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow. *Infotek: Jurnal Informatika Dan Teknologi*, 7(2), 500–510.
<https://doi.org/10.29408/jit.v7i2.26375>
- Suderajat, A., Kamalia, A. Z., & Afandi, D. (2025). Segmentasi Harga Pasar Properti Rumah menggunakan Algoritma K-Means berdasarkan Lokasi, Harga dan Luas Bangunan. *Dinamik*, 30(2), 180–192.
<https://doi.org/10.35315/dinamik.v30i2.10182>



Penulis pertama, Stefanus, merupakan mahasiswa Prodi Sistem Infomasi Universitas Putera Batam yang sangat menyukai bidang data analisis.



Penulis kedua, Darmansah, merupakan Dosen Prodi Sistem Informasi Universitas Putera Batam. Penulis banyak berkecimpung di bidang data analisis.