

BAB I

PENDAHULUAN

1.1 Latar Belakang

Rumah adalah salah satu dari kebutuhan sekunder yang penting untuk memilih sesuai dengan keinginan konsumen. Dalam proses pemilihan rumah terdapat beberapa kriteria seperti harga, lokasi, desain dan fasilitas rumah. Semua ini bertujuan untuk memastikan agar konsumen merasa puas dengan pilihan mereka[1].

Seiring dengan bertambahnya penduduk pemenuhan kebutuhan tempat tinggal tidak mudah dipenuhi, karena ada kriteria dan persyaratan pembangunan suatu rumah. Pelaksanaan pengembangan hunian dan perkampungan harus sesuai dengan peraturan tata ruang wilayah, sehingga dapat berjalan dengan teratur, terstruktur dengan baik sesuai dengan kebutuhan dan hukum yang berlaku[1].

Pemukiman merupakan bidang kajian yang sangat luas. Permasalahan yang berkaitan dengan factor ekonomi, geografis, sosial, budaya, demografik, politik, lingkungan, serta cara penanganan yang melibatkan berbagai pihak seperti pemerintah, arsitek, perencana, pengembang dan masyarakat sendiri[2]. Di beberapa wilayah dengan kondisi daerah maupun social ekonominya yang berbeda, permasalahan dalam penyelenggara rumah yang dihadapi berbeda pula. Misalnya masalah perumahan di wilayah pinggiran kota, perkotaan, dan pedesaan tidak akan sama.

Salah satu Teknik yang digunakan untuk mengidentifikasi suatu daerah berdasarkan beberapa variable yaitu metode klastering. Klastering merupakan suatu pendekatan statistik yang bertujuan untuk mengelompokkan pengamatan berdasarkan kemiripan karakteristik mereka dalam kelompok yang sama memiliki kesamaan yang lebih besar daripada pengamatan di kelompok lain. Teknik ini telah menarik minat dari berbagai kalangan dan telah dijelaskan dalam berbagai publikasi ilmiah[3].

Sebagai contoh [6] telah menganalisis kelompok penelitian di bidang sosial untuk pengumpulan status masyarakat dengan memperhatikan kesamaan dan perbedaan lintas budaya. Seiring itu, Wang et al melakukan studi pengembangan metode baru yang memungkinkan penggabungan dari beberapa data inframerah yang berasal dari dua jenis kanker yang berbeda. Mereka menemukan bahwa pendekatan ini dapat menghasilkan temuan yang signifikan dan berpotensi meningkatkan akurasi[4]. Algoritma umum yang digunakan dalam melakukan clustering ialah algoritma *K-Means*. Model pembelajaran ini digunakan untuk data yang belum pernah diklasifikasikan.

Klasterisasi bertujuan untuk mengelompokkan data ke dalam kelompok berdasarkan karakteristik atau fitur yang serupa. Dalam konteks penjualan rumah, klasterisasi dapat digunakan untuk mengidentifikasi segmen pasar yang berbeda berdasarkan preferensi pembeli, lokasi, atau karakteristik properti. Dengan mengelompokkan rumah yang memiliki kesamaan dalam faktor-faktor seperti harga, ukuran, dan fasilitas. Ini memungkinkan pemasar untuk menawarkan produk yang sesuai dengan kebutuhan dan keinginan konsumen dan meningkatkan peluang penjualan dan kepuasan pelanggan.

Selain itu, klasterisasi *K-Means* membantu dalam analisis tren pasar properti. Dengan mengidentifikasi pola dan kecenderungan dalam penjualan rumah, misalnya jika klasterisasi menunjukkan peningkatan minat terhadap properti di daerah pinggiran kota, maka penjualan akan berfokus pada pengembangan proyek baru di area tersebut.

K-means Clustering adalah data mining yang sering digunakan mengelompokkan menjadi beberapa klaster, di mana setiap klaster berisi data yang memiliki kemiripan sehingga ditempatkan dalam satu kelompok, sementara data yang tidak serupa akan ditempatkan dalam klaster yang berbeda, hal ini menyebabkan setiap klaster memiliki variasi data. Peran metode pengelompokan semakin meningkat seiring dengan berkembangnya teknologi dan kebutuhan manusia yang harus dipenuhi. Kebutuhan ini mencakup berbagai bidang seperti pendidikan, pemerintahan, pemasaran bisnis, dan pariwisata[5]. Salah satu teknik analisis pengelompokan yang umum digunakan adalah metode K-means.

Metode *K-Means* melakukan pendekatan analisis untuk pengelompokan objek ke dalam kluster melalui nilai rata-rata terdekat. Namun, untuk menilai kualitas hasil klusterisasi diperlukan metode evaluasi yang dapat mengukur seberapa baik kelompok yang terbentuk dan seberapa terpisahnya antar kelompok. Salah satu metode evaluasi yang digunakan adalah *Elbow* dan *silhouette coefficient* untuk mengukur kualitas dan keefektifan kluster[6].

K-Means memiliki beberapa keunggulan, termasuk kemudahan relatif dan kesederhanaannya dalam penerapan pada skala besar. *K-Means* juga dapat dengan mudah menyesuaikan diri dengan sampel baru yang diintegrasikan ke dalam cluster dengan berbagai bentuk dan ukuran, termasuk cluster yang berbentuk *elips*.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Metode *K-Means* memiliki beberapa kelemahan, yaitu penentuan jumlah kluster (*K*) yang optimal sering menjadi tantangan, namun hal ini dapat diatasi dengan metode *Elbow*, *Silhouette Score* untuk memilih jumlah kluster yang paling tepat. *K-Means* sangat sensitif terhadap outlier, yang dapat diatasi dengan preprocessing data. Untuk menguji kinerjanya, metode ini akan diterapkan pada data yang memiliki sifat linier pada data penjualan rumah. Proses yang memberikan hasil optimal akan digunakan untuk mengelompokkan kluster rumah berdasarkan tahun, luas bangunan, luas lahan, kamar tidur, kamar mandi, lokasi dan harga.

1.2.3 Solusi Permasalahan

Solusi permasalahan ini melibatkan Penerapan metode *K-Means* untuk melakukan pengelompokan data penjualan rumah di Kabupaten Sumenep. *K-means* digunakan untuk mengurutkan centroid dalam penentuan centroid awal, sehingga centroid yang dihasilkan bisa optimal. Sehingga nantinya didapatkan jumlah cluster paling optimal pada data tersebut.

1.2.2 Pertanyaan Penelitian

1. Bagaimana hasil penerapan metode *K-Means* untuk *clusterisasi* penjualan rumah kabupaten sumenep?
2. Bagaimana cara menentukan *k* terbaik pada metode *K-Means* dengan menggunakan optimasi *elbow*?

1.3 Batasan Masalah

1. Menggunakan metode *K-Means* dalam pengelompokan penjualan rumah di kabupaten sumenep.
2. Menggunakan Elbow dan silhouette Coefficient sebagai perbandingan evaluasi untuk menilai cluster yang terbaik dalam analisis data.
3. Data yang digunakan terdiri dari 319 data rumah yang terdapat di kabupaten sumenep.
4. Menggunakan 7 parameter yaitu Tahun, harga, lokasi, tempat tidur, kamar mandi, luas lahan, luas bangunan.

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

Tujuan dari penelitian ini sebagai berikut :

1. Untuk mengetahui hasil penerapan metode *K-Means* pada *clusterisasi* penjualan rumah Kabupaten Sumenep?
2. Untuk mengetahui K terbaik dari algoritma k-means menggunakan Sum of Squared Errors (SSE).

1.4.2 Manfaat

Manfaat yang akan di dapat pada penelitian ini yaitu :

1. Memberikan pemahaman yang lebih baik tentang karakteristik dan pola penjualan rumah.
2. Membantu pemerintah dan Lembaga terkait dalam penyusunan kebijakan penjualan rumah di kabupaten sumenep.
3. Memberikan kontribusi dalam pengembangan penelitian dan studi lebih lanjut dalam bidang clusterisasi penjualan rumah menggunakan metode *K-Means*.

1.5 Sistematika Laporan

Adapun sistematika penulisan proposal skripsi ini terbagi menjadi 3 bagian utama, yaitu :

1. BAB I PENDAHULUAN

Berisi tentang pendahuluan yang dilakukan pada penelitian ini. Bab ini membahas latar belakang permasalahan yang ada, solusi yang diusulkan untuk menyelesaikan masalah, tujuan dan manfaat yang dilakukan di penelitian, Batasan masalah serta sistematika penulisan proposal.

2. BAB II KAJIAN PUSTAKA

Bab ini membahas tentang kajian mengenai materi dan metode yang menjadi landasan teori dalam penelitian ini. Beberapa materi yang dibahas dalam bab ini yaitu data mining, *clustering*, algoritma *K-Means*, evaluasi hasil *clustering*, dan penelitian terkait algoritma *K-Means*.

3. BAB III METODOLOGI PENELITIAN

Bab ini membahas mengenai metode usulan yang ditawarkan oleh peneliti. Bab ini berisi penjelasan tentang laur penelitian, dataset, perhitungan manual algoritma *K-Means*.

4. BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan mengenai hasil dan pembahasan dari penelitian yang dilakukan.

5. BAB V PENUTUP

Bab ini membahas mengenai hasil kesimpulan dan saran dari peneliti.