

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan data penjualan rumah guna memberikan rekomendasi yang tepat bagi pengembang properti dalam meningkatkan strategi pemasaran. Algoritma K-Means Clustering digunakan sebagai metode utama untuk mengelompokkan data, dengan metode elbow membantu menentukan jumlah cluster optimal. Proses penelitian mengikuti metodologi CRISP-DM, mulai dari pemahaman bisnis hingga implementasi hasil. Data penjualan rumah diproses dengan normalisasi min-max dan label encoding. Algoritma K-Means menghitung jarak menggunakan euclidean distance dan mengelompokkan data berdasarkan atribut seperti harga rumah, luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, lokasi dan tahun. Dengan menggunakan algoritma *K-Means*, data tersebut dikelompokkan menjadi beberapa *cluster* berdasarkan pola yang terbentuk dari variabel-variabel tersebut. Hasil dari metode elbow menunjukkan jumlah optimal *cluster* adalah 3, yang di uji menggunakan optimasi elbow dan silhouette coefficient dan menghasilkan 3 klaster lah yang paling optimal dengan hasil SSE (Sum of Square Error) 25.9539. Dan hasil Silhouette Coefficient score 0.5215. Hasil *clustering* menghasilkan 3 *cluster* dengan jumlah per klaster yaitu, K=1 berjumlah 43 data, K=2 berjumlah 159 data, K=3 berjumlah 117 data.

Kata Kunci : Penjualan rumah, Data Mining, K-Means, *Clustering*

ABSTRACT

This research aims to cluster house sales data to provide accurate recommendations for property developers in enhancing their marketing strategies. The K-Means Clustering algorithm is used as the primary method for data clustering, with the elbow method helping to determine the optimal number of clusters. The research process follows the CRISP-DM methodology, starting from business understanding to the implementation of results. House sales data is processed using min-max normalization and label encoding. The K-Means algorithm calculates distances using Euclidean distance and clusters the data based on attributes such as house price, land area, building area, number of bedrooms, number of bathrooms, location, and year. Using the K-Means algorithm, the data is grouped into several clusters based on the patterns formed by these variables. The results of the elbow method show that the optimal number of clusters is 3, which was tested using elbow optimization and silhouette coefficient, producing the most optimal result of 3 clusters with an SSE (Sum of Square Error) of 25.9539. The silhouette coefficient score is 0.5215. The clustering results produced 3 clusters with the number of data points per cluster being: K=1 with 43 data points, K=2 with 159 data points, and K=3 with 117 data points.

Keywords: House sales, Data Mining, *K-Means*, *Clustering*