

## 1. Pendahuluan

### Latar Belakang

Rumah merupakan salah satu kebutuhan yang sangat penting bagi manusia dan termasuk ke dalam tiga kebutuhan primer manusia yaitu sandang, pangan, dan papan. Rumah merupakan tempat tinggal sekaligus tempat berlindung bagi manusia. Seiring berkembangnya zaman, kebutuhan manusia untuk memiliki rumah juga semakin tinggi, begitu juga dengan harga rumah yang semakin bervariasi bergantung pada parameter yang dimiliki tiap rumah seperti luas rumah, daerah tempat rumah, jumlah ruangan, dll.

Karena harga rumah yang semakin bervariasi, perlu dilakukan penelitian lebih lanjut untuk menganalisis keterkaitan antara harga rumah terhadap parameter-parameter lain yang dimiliki rumah tersebut. Analisis yang dilakukan diharapkan dapat mengelompokkan rumah ke dalam beberapa kategori sehingga orang-orang yang ingin membeli rumah dapat memiliki wawasan yang cukup dalam membeli rumah yang sesuai dengan kebutuhan dan anggaran yang mereka miliki. Sebelumnya sudah ada yang melakukan penelitian tentang segmentasi dan penentu harga rumah menggunakan metode *hedonic regression* [1].

*Clustering* merupakan salah satu proses unsupervised learning yang sangat penting dalam machine learning. *Clustering* merupakan suatu proses pengelompokan sekumpulan data yang “mirip” dalam satu *cluster* yang sama dan berbeda jika dibandingkan pada data yang berada dalam *cluster* yang lain [2]. *Clustering* biasanya digunakan dalam proses segmentasi suatu data, sehingga *clustering* juga sangat cocok untuk digunakan dalam penelitian ini.

*Hierarchical clustering* merupakan metode *clustering* pertama yang pernah digunakan yang digunakan oleh ahli biologi dan ilmuwan sosial, dimana pada saat itu analisis *cluster* telah menjadi cabang dari statistik analisis multivariat [3]. Sedangkan pada penelitian ini akan digunakan metode *Gaussian mixture model-based clustering*. *Gaussian mixture model* (GMM) merupakan sebuah model yang cukup sederhana untuk permasalahan *classification* maupun *clustering* jika dibandingkan dengan metode lain [4]. Pada akhir-akhir ini sudah banyak penelitian yang menggunakan GMM sebagai metode *clustering*. Sebagai contoh ada penelitian yang menggunakan *Gaussian mixture model-based clustering* untuk mengidentifikasi profil penggunaan listrik harian [5]. Hasil dari penelitian tersebut adalah GMM-based *clustering* dapat menghasilkan informasi berguna yang berkaitan dengan pola profil listrik harian dari sebuah gedung. Selain itu, penelitian ini juga menyatakan kalau GMM-based *clustering* menghasilkan biaya komputasi yang lebih rendah jika dibandingkan dengan *hierarchical clustering*. **Tabel 1** menjelaskan perbandingan antara GMM dan *hierarchical clustering* menggunakan metode *silhouette* dari penelitian yang mengidentifikasi profil penggunaan listrik harian.

**Tabel 1. Perbandingan GMM-based clustering dengan hierarchical clustering dalam mengidentifikasi profil penggunaan listrik harian**

Metode <i>clustering</i>	Jumlah <i>cluster</i> optimal	Nilai skor <i>silhouette</i>
GMM	2	0,719
<i>Hierarchical clustering</i>	2	0,696

Kedua metode, baik GMM maupun *hierarchical clustering* memiliki jumlah *cluster* optimal yang sama yaitu sebanyak dua *cluster*. Namun jika dilihat skor *silhouette*-nya, GMM menghasilkan skor yang lebih besar jika dibandingkan dengan *hierarchical clustering*.

Selain itu, terdapat penelitian lain yang membandingkan antara metode K-Means dengan GMM[6]. Dataset yang digunakan adalah dataset dari Google Cluster Trace dengan masalah utamanya adalah tentang kualitas *cluster* dan waktu komputasinya. **Tabel 2** merupakan hasil dari penelitian tersebut.

**Tabel 2. Perbandingan GMM-based clustering dengan K-Means clustering dengan menggunakan Google Cluster Trace dataset**

Metode <i>clustering</i>	Jumlah <i>cluster</i> optimal	Waktu komputasi (ms)
GMM	13	312.03
K-Means	2	46.86

Kedua metode memiliki perbedaan yang cukup signifikan. GMM memiliki jumlah *cluster* optimal yang jauh lebih banyak dibandingkan dengan K-Means. Hal ini menunjukkan kalau metode GMM dapat memperoleh informasi yang lebih spesifik jika dibandingkan dengan K-Means. Karena lebih banyak *cluster* berarti lebih banyak informasi-informasi kecil yang dapat diperoleh. Walaupun GMM memiliki waktu komputasi yang lebih lama dibandingkan K-Means, namun hal ini dapat dimaklumi karena jumlah *cluster* optimal pada GMM jauh lebih banyak dibandingkan K-Means.

**Topik dan Batasannya**

- Bagaimana cara melakukan optimasi parameter dari GMM?
- Bagaimana hasil segmentasi harga rumah menggunakan *GMM1-Based Clustering* dan informasi apa saja yang bisa didapat?

**Tujuan**

- Mengidentifikasi cara melakukan optimasi parameter dari GMM
- Mengidentifikasi hasil segmentasi harga jual rumah menggunakan *GMM-Based Clustering* dan menganalisis hasil yang didapatkan.