

# BAB I PENDAHULUAN

## 1.1. Latar Belakang

Rumah merupakan salah satu bentuk investasi yang menarik. Permintaan akan tempat tinggal di Kabupaten Bogor masih terus meningkat akibat pertumbuhan penduduk akibat urbanisasi dan migrasi dari luar kota. Pertumbuhan penduduk dan kepadatan penduduk adalah dua faktor yang mempengaruhi pembangunan perumahan di Kabupaten Bogor, dan sekali lagi semua harga menjadi masalah. Perumahan dan tempat tinggal juga merupakan kebutuhan dasar manusia. Selain itu, rumah juga merupakan kebutuhan dasar manusia untuk perbaikan martabat, kualitas hidup, penghidupan. Dan sebagai refleksi dalam upaya meningkatkan taraf hidup dan pembentukan kepribadian kebangsaan dan karakter.

Menurut data di atas, dengan melihat jumlah penduduk di Kabupaten Bogor yang semakin bertambah seiring waktu, masyarakat tentu membutuhkan tempat tinggal yang layak dan nyaman untuk ditinggali. Dengan laju pertumbuhan penduduk yang cukup cepat maka dibutuhkan banyak tempat tinggal yang disiapkan untuk beberapa tahun ke depan. Bersamaan dengan data kurs fluktuatif yang pada tahun 2018, 1 USD (United States Dollar) sama dengan 15.233 Rupiah, pada 2019, 1 USD (United States Dollar) sama dengan 14.212 Rupiah, dan pada 2020, 1 USD (United States Dollar) sama dengan 14.034 Rupiah.

Penelitian ini bertujuan untuk menguji performa dua model, yaitu *clustering* dengan menggunakan *K-Means* dan *Gaussian Mixture Model* dan akan dibandingkan hasil proses dari kedua metode tersebut. *Clustering* merupakan *proses machine learning* yang berfungsi untuk mengelompokkan sekumpulan data yang memiliki kesamaan karakteristik menjadi satu cluster. Sebagai contoh Pada penelitian sebelumnya, ada penelitian yang menggunakan *Gaussian Mixture Model* untuk mengidentifikasi kepadatan kendaraan di jalan tol, dan menghasilkan tingkat akurasi sebesar 91% [15]. Lalu ada identifikasi area kanker ovarium pada citra CT Scan, yang memperoleh hasil tidak cukup bagus dengan persentasi *true positive* sebesar 45% lebih kecil dari *false positive* sebesar 55% [16]. Hasil dari penelitian

tersebut memberi informasi tentang prediksi jumlah total kasus COVID-19 di seluruh dunia dan prediksi tanggal berakhirnya pandemic COVID-19. Lalu terdapat penelitian yang melakukan perbandingan *clustering cloud workloads* dengan menggunakan dua metode yaitu *K-Means* dan *Gaussian Mixture Model*[11]. Dari penelitian tersebut menghasilkan *Gaussian Mixture Model* memberikan *cluster* yang lebih baik, dan lebih rinci daripada *K-Means*, namun dibutuhkan waktu yang lebih lama dibanding proses *clustering* pada *K-Means*.

## 1.2. Perumusan Masalah<sup>[1][2]</sup>

Rumusan masalah yang dihadapi dalam proposal tugas akhir ini secara umum :

1. Bagaimana performa dan kualitas clustering harga rumah menggunakan *K-Means* dan *GMM*?

## 1.3. Tujuan

Berdasarkan rumusan masalah, tujuan penelitian Tugas Akhir ini adalah sebagai berikut :

1. Menguji performa dan kualitas hasil clustering harga rumah dengan menggunakan *Gaussian Mixture Model* dan *K-Means*.

## 1.4. Batasan Masalah

Adapun batasan masalah yang menjadi lingkup pada tugas akhir ini adalah sebagai berikut:

1. Data rumah yang dipakai pada penelitian kali ini adalah data rumah di Kabupaten Bogor yang dihimpun dari website [olx.co.id](http://olx.co.id)
2. Atribut yang digunakan untuk penelitian adalah harga rumah, karena tujuan penelitian ini hanya menguji performansi dari dua model, bukan untuk membuat informasi data rumah secara detail.
3. Pembanding yang digunakan adalah nilai *Silhouette Score*.

## 1.5. Metode Penelitian

Penelitian ini akan dilakukan dengan kegiatan sebagai berikut:

1. Bimbingan dengan dosen pembimbing  
Kegiatan bimbingan dengan dosen pembimbing dilakukan secara daring kepada dosen-dosen pembimbing untuk mendiskusikan hal yang berkaitan dengan penelitian ini.
2. Studi Literatur  
Kegiatan ini meliputi pengumpulan literatur terkait yang berhubungan dengan tugas akhir ini. Beberapa diantaranya mengenai ***Big data, machine learning, K-Means, dan Gaussian Mixture Model*** dari jurnal, buku, dan penelitian-penelitian sebelumnya.
3. Pengumpulan dan Pra-Processing Data  
Melakukan pengumpulan data spesifikasi dan harga rumah melalui proses *scraping* pada *website* olx.co.id. Lalu data yang telah dikumpulkan melalui proses *scraping* akan dilakukan proses penghapusan variabel-variabel yang tidak digunakan.
4. Proses Clustering  
Pada tahap ini dilakukan proses clustering pada dataset yang telah dibuat, dengan menggunakan model K-Means dan dibandingkan dengan Gaussian Mixture Model.
5. Analisis Perbandingan Hasil Clustering  
Setelah dilakukan proses clustering menggunakan model K-Means, selanjutnya akan dilakukan perbandingan hasil clustering yang menggunakan model K-means dengan hasil clustering menggunakan Gaussian Mixture Model.
6. Penulisan Laporan  
Setelah melakukan tahap-tahap pengujian, lalu dilakukan penyusunan laporan tugas akhir sesuai sistematika penulisan yang telah ditetapkan.