

# BAB I

## PENDAHULUAN

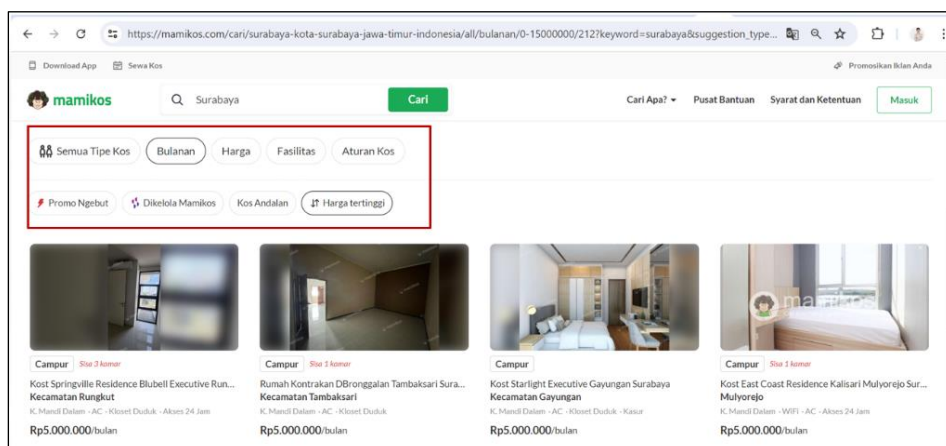
### 1.1. Latar Belakang

Surabaya merupakan kota metropolis terbesar kedua setelah Jakarta dan menjadi salah satu kota tujuan urbanisasi (Astiyan Putra & Harianto, 2022). Kota Surabaya memegang peran penting dalam banyak aspek kehidupan nasional, termasuk dari aspek pendidikan. Surabaya menjadi kota yang dikenal maju dalam mendukung perkembangan intelektual yang dapat dilihat dari keberadaan banyak lembaga pendidikan dan universitas terkemuka di Kota Surabaya. Mengacu pada data BPS Jatim (2022), tercatat 70 perguruan tinggi di Kota Surabaya. Kondisi ini turut mendorong terjadinya migrasi pendidikan ke kota ini. Proses migrasi pendidikan terjadi karena para migran berkeinginan untuk menyambung studi mereka ke tingkat yang lebih tinggi dan mendapatkan pendidikan yang lebih berkualitas (Mustadi *et al.*, 2021). Setiap tahun, ribuan mahasiswa baru dari berbagai latar belakang bermigrasi ke kota-kota besar untuk melanjutkan pendidikan. Peningkatan dalam pertumbuhan penduduk perkotaan melalui migrasi (dalam hal ini mahasiswa sebagai pendatang) sejalan dengan permintaan yang tinggi akan tempat tinggal. Situasi ini menjadi alasan semakin banyak hunian sejenis indekos yang dapat ditemui di sekitar wilayah kota.

Pertumbuhan sekolah dan universitas berdampak pada meningkatnya kebutuhan properti indekos (Wibawa & Dewa, 2024). Saat memilih rumah kos atau indekos, mahasiswa kerap mempertimbangkan banyak hal. Pada penelitian Paruntu *et al.* (2021) disebutkan bahwa dalam menentukan pilihan indekos, mahasiswa sangat memperhatikan faktor lokasi dan harga. Selain itu, kelengkapan fasilitas yang tersedia juga tidak kalah pentingnya dijadikan pertimbangan memilih indekos (Wijoyo & Maimunah, 2019). Namun, dengan beragam pilihan dan fasilitas indekos di berbagai lokasi, mahasiswa sering kali menghadapi kesulitan dalam menentukan akomodasi yang paling sesuai dengan kebutuhan pribadi mereka.

Antar mahasiswa memiliki preferensi yang tidak sama dalam mempertimbangkan tempat tinggal atau indekos (Bartasari & Arnellis, 2023). Heterogenitas kriteria indekos seringkali menimbulkan tantangan tersendiri bagi calon penyewa untuk mengidentifikasi indekos yang sesuai dengan kebutuhan dan preferensi mereka (Sipayung *et al.*, 2021). Akibat hal tersebut, banyak mahasiswa sering salah memilih dan sering berpindah-pindah kos (Kusumantara *et al.*, 2020). Oleh karena itu, kondisi ini membutuhkan suatu pendekatan *data science* berupa sistem rekomendasi yang dapat membantu proses pencarian indekos yang ideal sesuai preferensi masing-masing mahasiswa.

Pendekatan yang saat ini digunakan oleh *platform* daring untuk menyediakan rekomendasi indekos, sering kali tidak memberikan hasil yang optimal karena masih menggunakan mekanisme *filtering* dasar. Sistem berbasis filter ini sering kali menghasilkan banyak pilihan, yang kemudian harus disaring lagi secara manual oleh pengguna apabila mempertimbangkan faktor tambahan lain seperti kedekatan dengan tempat-tempat tertentu dan sebagainya yang tidak bisa ditangani oleh sistem berbasis filter. Proses ini dapat memakan waktu dan membingungkan, terutama bagi pelajar atau mahasiswa yang baru pertama kali mencari indekos di kota besar seperti Surabaya. Keterbatasan lainnya dari pendekatan berbasis filter ini adalah bahwa *platform* tidak dapat menampilkan secara terurut indekos yang paling dekat dengan ekspektasi dan preferensi yang di-*input* oleh pengguna.



**Gambar 1.1** Sistem berbasis Filter pada Salah Satu *Platform* Indekos

Salah satu pendekatan *data science* yang dapat dimanfaatkan untuk memberikan rekomendasi indeks adalah analisis kluster. Prinsip dasar analisis kluster adalah mengelompokkan data (dalam hal ini indeks) ke dalam kelompok-kelompok di mana objek-objek dalam kelompok yang sama menunjukkan level kesamaan yang tinggi, sedangkan objek dalam kelompok lain akan terlihat relatif berbeda (Haryati & Sugiyarto, 2021). Ketika diterapkan pada konteks indeks, analisis kluster dapat digunakan untuk mengidentifikasi segmen atau pengelompokan rumah kos yang berbeda berdasarkan metrik kesamaan yang terdapat dalam data.

Adapun pemilihan metode analisis kluster juga bergantung pada jenis data yang digunakan. Untuk data numerik, metode seperti *K-Means* yang mengelompokkan data berdasarkan *Euclidean Distance* terbukti efektif (Fawaid Ridwan & Supian, 2021). Namun, dalam banyak penelitian metode *K-Means* tidak bekerja dengan baik pada atribut kategorikal (Harous *et al.*, 2019). Dalam kasus seperti ini, teknik seperti *K-Modes*, yang menggunakan modus dalam setiap kluster sebagai *centroid*, lebih cocok untuk menangani data *Non-numerik* (Llop Torrent, 2019). Lalu, untuk melakukan klusterisasi pada data yang mengandung campuran atribut numerik dan kategorik, *Gower Distance* sebagai metrik jarak dapat digunakan untuk menerapkan metode klusterisasi.

Metode klusterisasi *K-Prototype* dan *Partitioning Around Medoids* (PAM) dengan metrik jarak Gower pada data campuran terbukti bekerja dengan baik dalam kasus pengelompokan pada banyak penelitian. Penelitian yang dilakukan oleh Yildirim *et al.* (2019) berfokus pada pengelompokan pasien perawatan primer di sebuah rumah sakit di New York berdasarkan berbagai faktor kesehatan. Studi ini menggunakan metode pengelompokan *K-Prototype* untuk mendapatkan pemahaman terkait kebutuhan kesehatan setiap kelompok pasien. Hasil penelitian mengungkapkan bahwa populasi pasien dapat diklasifikasikan menjadi empat kelompok berbeda dengan masing-masing kelompok mewakili pasien dengan karakteristik kesehatan serupa. Hal ini memungkinkan rumah sakit dan pembuat kebijakan untuk memberikan perawatan yang lebih spesifik dan efektif yang disesuaikan

dengan kebutuhan unik setiap kelompok pasien. Penelitian selanjutnya oleh Asensio *et al.* (2022) yang melakukan pengelompokan calon mahasiswa ke dalam kelompok yang homogen untuk mengidentifikasi calon mahasiswa potensial untuk pendaftaran di suatu universitas. Peneliti menerapkan analisis klaster dengan metode *Partitioning Around Medoids* (PAM) yang dipadukan dengan *Gower Distance* untuk mengukur jarak antar kandidat yang terdiri dari perpaduan variabel data numerik dan kategorik. Pendekatan ini berhasil membentuk kelompok calon mahasiswa dengan karakteristik serupa sehingga memungkinkan universitas menyesuaikan layanan dan strategi pemasaran mereka untuk meningkatkan jumlah pendaftaran. Lebih lanjut, penelitian yang dilakukan oleh Fadilah & Wijayanto (2023) meneliti beberapa algoritma pengelompokan dalam studi kasus penyakit ginjal kronis. Peneliti membandingkan berbagai metode, antara lain *one-hot encoding*, metrik *Gower Distance* yang dipadukan dengan algoritma seperti *K-Means*, *DBSCAN*, *OPTICS*, *Partitioning Around Medoids* (PAM), dan *K-Prototype* untuk dataset dengan tipe data campuran. Berdasarkan evaluasi indeks *Silhouette*, peneliti menemukan bahwa *K-Prototype* dan *Partitioning Around Medoids* (PAM) dengan metrik *Gower Distance* merupakan metode yang paling optimal, mencapai indeks *Silhouette* tertinggi dibandingkan teknik lain yang digunakan dalam penelitian ini.

Mengingat latar belakang dan pengaruh faktor numerik dan kategorikal terhadap keputusan mahasiswa dalam memilih indekos seperti yang telah dijabarkan diatas, maka peneliti menggunakan teknik klasterisasi untuk mengelompokkan indekos dengan metode *K-Prototype* dan *Partitioning Around Medoids* (PAM) dengan metrik *Gower Distance*. Objek penelitian ini adalah indekos yang berlokasi di wilayah Surabaya, khususnya di sekitar kampus UPN “Veteran” Jawa Timur yang menjadi objek acuan utama. Dengan menerapkan teknik klasterisasi, peneliti bertujuan untuk mengidentifikasi kelompok-kelompok indekos yang memiliki karakteristik serupa. Urgensi penelitian ini didasarkan pada heterogenitas kriteria indekos seringkali menimbulkan tantangan tersendiri bagi calon penyewa yang pada akhirnya sering ditemukan permasalahan pada mahasiswa yakni sering

berpindah-pindah kos. Oleh karena itu, perlu dibangun sistem rekomendasi berbasis preferensi yang dapat membantu mahasiswa menemukan indekos yang sesuai dengan kebutuhan. Peneliti akan membandingkan kinerja kedua metode untuk menentukan pendekatan yang paling sesuai untuk memprediksi klaster yang selaras dengan masukan pengguna. Adapun kebaharuan dari penelitian ini adalah dalam memberikan rekomendasi didasarkan pada hasil klasterisasi dan dikombinasikan dengan perhitungan jarak antara *input* pengguna dengan masing-masing indekos dalam klaster yang sama dengan yang diprediksi. Sehingga akan memungkinkan identifikasi rumah kos yang paling dekat (dengan jarak terkecil) dan relevan dengan preferensi pengguna.

## 1.2. Rumusan Masalah

Mengacu pada penjeasan latar belakang diatas, dibawah ini adalah rumusan masalah yang akan dijawab dan menjadi fokus penelitian:

- 1) Bagaimana proses pemodelan klaster menggunakan metode *K-Prototype* dan *Partitioning Around Medoids* (PAM) dengan metrik *Gower Distance*?
- 2) Bagaimana performa hasil pengelompokan indekos serta karakteristik dari masing-masing klaster indekos yang terbentuk?
- 3) Bagaimana proses sistem dalam memberikan rekomendasi indekos yang paling sesuai dengan preferensi pengguna menggunakan kombinasi hasil klasterisasi dan perhitungan *disimilarity* dengan *Gower Distance*?

## 1.3. Batasan Masalah

Dalam konteks riset ini, berikut merupakan batasan yang perlu diperhatikan:

- 1) *Scope* data indekos diambil dari *platform* Mamikos
- 2) Dalam proses *scrapping* data indekos, peneliti berfokus pada indekos yang berada di sekitar wilayah Surabaya Timur khususnya wilayah sekitar kampus UPN “Veteran” Jawa Timur

#### 1.4. Tujuan Penelitian

Berlandaskan rumusan masalah yang telah dijelaskan, tujuan utama yang ingin dicapai melalui riset ini adalah sebagai berikut:

- 1) Melakukan proses pemodelan klaster menggunakan metode *K-Prototype* dan *Partitioning Around Medoids* (PAM) dengan metrik *Gower Distance*
- 2) Membandingkan performa hasil pengelompokan indekos dan mengidentifikasi karakteristik dari masing-masing klaster indekos yang terbentuk.
- 3) Melakukan *deployment* terkait dengan proses pemberian rekomendasi indekos yang paling sesuai dengan preferensi pengguna melalui kombinasi hasil klasterisasi dan perhitungan *Gower Distance*.

#### 1.5. Manfaat Penelitian

Mengacu pada tujuan penelitian yang diuraikan diatas, dibawah ini merupakan manfaat teoritis dan praktis dari penelitian ini:

##### 1) Manfaat Teoritis

Penelitian ini diharapkan dapat memperkaya literatur ilmiah dengan berkontribusi pada pemahaman baru terkait metode klasterisasi dengan metrik jarak Gower yang digunakan dalam membangun sistem rekomendasi.

##### 2) Manfaat Praktis

Output penelitian dapat dapat dijadikan panduan oleh para pemangku kepentingan, terutama bagi mahasiswa dalam mencari indekos di sekitar wilayah kampus UPN “Veteran” Jawa Timur.