

Analisis Klaster *Partitioning Around Medoids* dengan *Gower Distance* untuk Rekomendasi Indekos (Studi Kasus: Indekos di Sekitar Kampus UPNVJT)

Sahat Renaldi. S^{1✉}, Dwi Arman Prasetya², Amri Muhaimin³

^{1, 2, 3} Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur, Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 04-07-2024

Direvisi : 09-07-2024

Diterima : 12-07-2024

Kata Kunci:

Gower Distance; Indekos; K-Prototype; *Partitioning Around Medoids* (PAM); Sistem Rekomendasi.

Keywords :

Gower Distance; Boarding House; K-Prototype; *Partitioning Around Medoids* (PAM); Recommendation System.

Corresponding Author :

Sahat Renaldi. S

Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur
Jl. Rungkut Madya No. 1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur, 60294

Email: srenaldd@gmail.com

ABSTRAK

Dalam memilih indekos, ada banyak kriteria yang dipertimbangkan calon penyewa. Kondisi ini membutuhkan pendekatan analitik yang dapat membantu individu dalam mencari indekos dengan tetap memperhatikan preferensi masing-masing. Penelitian ini bertujuan membangun sistem rekomendasi melalui pendekatan analisis kluster K-Prototype dan *Partitioning Around Medoids* (PAM) dengan *Gower Distance*. Metode PAM dengan *Gower Distance* bekerja lebih baik dengan indeks *Silhouette* 0.45 dan berhasil mengelompokkan indekos kedalam 2 kluster, yakni indekos standar dan eksklusif. Sistem rekomendasi kemudian dibangun berdasarkan kesamaan jarak Gower. Hasil pengujian dengan *User Acceptance Test* (UAT) sebesar 70% mengindikasikan bahwa secara keseluruhan sistem dapat dikatakan baik dan dapat diterima oleh pengguna.

ABSTRACT

In choosing a boarding house, many criteria are considered by prospective tenants. This situation requires an analytical approach that can assist individuals in finding a boarding house while taking into account their preferences. This study aims to develop a recommendation system using K-Prototype and Partitioning Around Medoids (PAM) with Gower Distance. The PAM method with Gower Distance performed better, achieving a Silhouette index of 0.45, and successfully clustered boarding houses into two categories: standard and exclusive. The recommendation system is based on Gower distance similarity. User Acceptance Test (UAT) results showed a score of 70%, indicating that the system is generally well-received and acceptable to users.

PENDAHULUAN

Surabaya merupakan kota metropolitan terbesar kedua setelah Jakarta dan menjadi salah satu kota tujuan urbanisasi (Astiyan Putra & Harianto, 2022). Surabaya memegang peran penting dalam banyak aspek kehidupan nasional, termasuk dari aspek pendidikan yang bisa dilihat dari banyaknya keberadaan lembaga pendidikan dan universitas di kota ini. Hal ini tentunya turut mendorong terjadinya migrasi pendidikan. Pertumbuhan sekolah dan universitas berdampak pada meningkatnya kebutuhan akan 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, antar mahasiswa yang satu dengan yang lain memiliki preferensi yang berbeda dalam memilih tempat tinggal atau indekos (Bartasari & Arnellis, 2023). Selain itu, keberagaman kriteria indekos kerap menimbulkan tantangan tersendiri bagi calon penyewa dalam mencari indekos yang sesuai dengan kebutuhan dan preferensi mereka (Sipayung *et al.*, 2021). Akibatnya, banyak mahasiswa sering salah pilih kos dan berujung pada kasus sering pindah-pindah kos (Kusumantara *et al.*, 2020). Oleh karena itu, kondisi ini membutuhkan suatu pendekatan analitik yang dapat membantu mahasiswa dalam proses pencarian indekos yang ideal dengan tetap memperhatikan preferensi dari masing-masing individu. Salah satu pendekatan analitik yang dapat dimanfaatkan untuk memberikan rekomendasi indekos adalah analisis kluster.

Prinsip utama analisis kluster adalah mengelompokkan data dimana data dalam kelompok yang sama akan memiliki kemiripan yang tinggi, sedangkan objek lain dalam kluster yang berbeda akan terlihat relatif berbeda (Haryati & Sugiyarto, 2021). Untuk data campuran seperti yang digunakan dalam ini, metode K-Prototype dan Partitioning Around Medoids (PAM) dengan Gower *Distance* sebagai metrik jarak digunakan untuk menghasilkan kluster yang optimal. Metode analisis kluster K-Prototype dan Partitioning Around Medoids (PAM) dengan Gower Distance untuk data campuran terbukti bekerja dengan baik dalam banyak penelitian. Salah satunya dalam penelitian yang dilakukan oleh Fadilah & Wijayanto (2023) meneliti beberapa algoritma pengelompokan dalam studi kasus penyakit ginjal kronis dengan data campuran. Peneliti membandingkan berbagai metode, antara lain *one-hot encoding*, Gower *Distance* yang dipadukan dengan algoritma seperti K-Means, DBSCAN, OPTICS, Partitioning Around Medoids (PAM), dan K-Prototype. Berdasarkan evaluasi indeks *silhouette*, peneliti menemukan bahwa K-Prototype dan Partitioning Around Medoids (PAM) dengan Gower *Distance* merupakan metode yang paling optimal, mencapai indeks *silhouette* tertinggi dibandingkan metode lain.

Berdasarkan uraian diatas, peneliti menggunakan teknik klusterisasi untuk mengelompokkan indekos dengan menggunakan metode K-Prototype dan Partitioning Around Medoids (PAM) dengan Gower *Distance*. Objek penelitian ini adalah indekos yang berlokasi di wilayah Surabaya, khususnya di sekitar kampus UPN “Veteran” Jawa Timur yang menjadi patokan peneliti. Dalam prosesnya, akan dilakukan perbandingan kinerja kedua metode ini untuk menentukan pendekatan yang paling sesuai. Selanjutnya dengan model kluster tersebut, akan dihitung jarak antara *input* pengguna dengan masing-masing indekos dalam kluster yang sama dengan yang diprediksi. Sehingga nantinya akan memungkinkan identifikasi rumah kos yang paling dekat (dengan jarak terkecil) dari preferensi pengguna. Kemudian sistem dikembangkan ke dalam aplikasi web menggunakan R Shiny. Peneliti menguji kebaikan sistem yang telah dibangun dengan melakukan *User Acceptance Test* (UAT).

METODE PENELITIAN

Pengumpulan dan Pembersihan Data

Proses pengumpulan data dilakukan menggunakan pendekatan hybrid, yaitu memanfaatkan teknik *web scraping* dan ekstraksi manual dari situs Mamikos. Peneliti melakukan pengambilan data secara manual pada data alamat indekos yang berupa link Google Maps. Link Google maps ini nantinya akan diekstraksi ke dalam informasi latitude dan longitude untuk kemudian dihitung jarak antara indekos ke kampus UPN “Veteran” Jawa Timur dengan menggunakan konsep *haversine*. Rumus Haversine digunakan untuk mencari jarak garis lurus antara dua koordinat melalui latitude dan longitude (Prasetya *et al.*, 2020). Selanjutnya, pembersihan data mencakup beberapa langkah yang bertujuan untuk mempersiapkan data dan memastikan kualitas data sebelum dilakukan analisis lanjutan.

Exploratory Data Analysis (EDA) dan Pra-Pemodelan

Dalam tahap ini, pendekatan yang digunakan didasarkan atas metode statistika deskriptif untuk memberikan gambaran umum mengenai karakteristik data yang telah dikumpulkan. Sebelum dilakukan pemodelan, penting untuk memastikan bahwa data yang akan digunakan

sudah sesuai dengan persyaratan dan standar yang diperlukan. Tahap ini mencakup pengecekan anomali. Anomali juga dapat disebut outlier, yakni objek data yang tampak berbeda secara signifikan dibandingkan dengan yang lain (Muhaimin, 2018). Selain itu, dilakukan juga pengecekan frekuensi kategori dalam suatu variabel untuk memastikan tidak ada ketimpangan data yang nantinya akan mempengaruhi hasil klusterisasi. Selanjutnya, dilakukan proses standarisasi pada variabel numerik untuk memastikan bahwa data memiliki skala yang seragam.

Pemodelan

Pada tahap ini akan dilakukan pemodelan kluster menggunakan metode K-Prototype dan Partitioning Around Medoids (PAM) dengan menggunakan Gower *Distance* sebagai metrik jarak. K-Prototype adalah metode klusterisasi yang menggabungkan algoritma K-Means untuk data numerik dengan algoritma K-Modes untuk data kategorikal. Cara kerja algoritma K-Prototype adalah sebagai berikut:

- Pilih k prototipe awal dan inisialisasi *centroid* untuk setiap kluster.
- Untuk setiap titik data, hitung jarak ke setiap *centroid*/prototipe. Berikut adalah formulanya (Nooraeni et al., 2019):

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2 + \gamma \sum_{i=1}^n \delta(x_i, y_i)} \quad (1)$$

dengan,

x dan y : dua vektor data yang memiliki n dimensi

x_i dan y_i : masing-masing merupakan koordinat atau nilai dari atribut ke-
i dari vektor x dan y

$\delta(x_i, y_i)$: menyatakan atribut yang cocok

γ : parameter bobot untuk atribut kategori pada kelompok i

- Untuk setiap kluster, hitung ulang centroid dengan menggunakan *mean* untuk atribut numerik dan modus untuk atribut kategorik. Proses pembaruan kluster dan centroid diulang sampai konvergensi tercapai.

Partitioning Around Medoids (PAM) adalah algoritma klusterisasi yang bertujuan untuk mempartisi data ke dalam sejumlah kluster yang telah ditentukan sebelumnya dengan menggunakan medoid. Rumus yang digunakan untuk menghitung total *cost* C dalam PAM adalah:

$$d(x, y) = \sum_{i=1}^n \min_j d(x_i, m_j) \quad (2)$$

dengan,

m_j : medoid ke- j dari sekumpulan medoid yang telah dipilih

$d(x_i, m_j)$: jarak antara titik data x_i dengan medoid m_j , dalam penelitian ini
menggunakan jarak Gower

$\min_j d(x_i, m_j)$: jarak minimum antara titik data x_i dengan medoid m_j

Gower *Distance* adalah sebuah metrik kesamaan yang digunakan untuk mengukur jarak antara dua entitas dalam data yang memiliki tipe variabel campuran. Jarak gower dapat dihitung dengan persamaan berikut (Prasath et al., 2017):

Untuk data nominal:

$$d_k(x_{ik}, y_{jk}) = \begin{cases} 0, & \text{saat } x_{ik} = y_{jk} \\ 1, & \text{saat } x_{ik} \neq y_{jk} \end{cases} \quad (3)$$

Untuk data ordinal:

$$d_k(x_{ik}, y_{jk}) = \frac{|r_k(x_{ik}) - r_k(x_{jk})|}{\max\{r_k(x_{\cdot k})\} - \min\{r_k(x_{\cdot k})\}} \quad (4)$$

Untuk data numerik:

$$d_k(x_{ik}, y_{jk}) = \frac{|(x_{ik}) - (x_{jk})|}{\max\{(x_{\cdot k})\} - \min\{(x_{\cdot k})\}} \quad (5)$$

dengan,

$x_{\cdot k}$: nilai seluruh observasi variabel k

$r_k(x_{\cdot k})$: rank seluruh observasi variabel ordinal k

$\max\{(x_{\cdot k})\}$: nilai maksimum dari seluruh variabel k

$\min\{(x_{\cdot k})\}$: nilai minimum dari seluruh variabel k

$\max\{r_k(x_{\cdot k})\}$: rank maksimum dari seluruh variabel ordinal k

$\min\{r_k(x_{\cdot k})\}$: Rank minimum dari seluruh variabel ordinal k

Untuk menghitung jarak antara dua objek digunakan persamaan berikut:

$$d(x, y) = \frac{\sum_{k=1}^p \delta_k(x_{ik}, y_{jk}) d_k(x_{ik}, y_{jk})}{\sum_{k=1}^p \delta_k(x_{ik}, y_{jk})} \quad (6)$$

dengan $\delta_k(x_{ik}, y_{jk}) = \begin{cases} 1, & \text{saat } x_{ik} \text{ atau } y_{jk} \text{ tidak hilang} \\ 0, & \text{lainnya} \end{cases}$

Evaluasi

Dalam menentukan jumlah kluster optimum dan mengevaluasi kebaikan model akan digunakan metode Elbow dan *Silhouette*. Metode Elbow merupakan metode yang melibatkan pengamatan nilai SSE (*Sum of Squared Errors*) untuk setiap jumlah kluster yang diuji (Ashari *et al.*, 2023). Langkah awal metode ini melibatkan pemilihan sejumlah nilai kluster yang berbeda untuk digunakan dalam pembentukan model data. Kemudian, dilakukan perhitungan persentase variasi data yang dijelaskan oleh setiap jumlah kluster tersebut. Hasil persentase tersebut dianalisis dengan membentuk grafik, dan jika terdapat siku yang mencirikan penurunan persentase yang paling signifikan antara dua nilai kluster berurutan, maka jumlah kluster yang terkait dengan siku tersebut dianggap sebagai jumlah kluster optimal. Metode *Silhouette* pada dasarnya mengevaluasi seberapa baik suatu objek terintegrasi dalam kluster tempatnya berada. Dewa & Jatipaningrum (2019) menyatakan bahwa untuk setiap objek dalam suatu kluster, *Silhouette Coefficient* menghitung rata-rata jarak antara objek tersebut dengan semua objek dalam kluster yang sama (*cohesion*), serta menghitung rata-rata jarak antara objek tersebut dengan objek-objek dalam kluster terdekat lainnya (*separation*).

HASIL DAN PEMBAHASAN

Pengumpulan dan Pembersihan Data

Data yang dikumpulkan fokus pada area Surabaya Timur. Setelah data diperoleh, kemudian dilakukan pembersihan data hingga pada akhirnya menghasilkan 677 entri data lengkap yang akan digunakan dalam proses eksplorasi. Berikut adalah sampel data yang telah berhasil dibersihkan:

Tabel 1. Data Bersih

No	Lokasi	Harga	Jenis	Listrik	Luas	K. Mandi Dalam	AC	WiFi	Kasur	Akses 24 Jam	Jarak ke UPN
1	Gubeg	1200000	Putra	Ya	9.0	Yes	No	Yes	Yes	Yes	1.2
2	Gubeg	1250000	Putri	Ya	12.5	Yes	Yes	Yes	Yes	Yes	2.3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
677	Gubeg	900000	Campur	Ya	8.0	No	No	Yes	Yes	Yes	3

Exploratory Data Analysis (EDA) dan Pra-Pemodelan

Berikut ini adalah ringkasan / statistik deskriptif dari sebaran data Harga, Luas, dan Jarak ke kampus UPN:

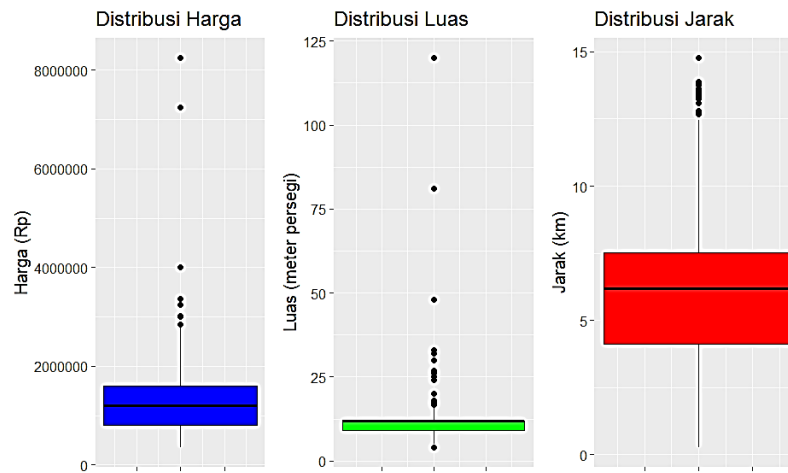
Tabel 2. Statistik Deskriptif Variabel Numerik

Statistik Deskriptif	Harga (rp)	Luas (meter ²)	Jarak (km)
Minimum	350000	4.0	0.26
Kuartil 1	800000	9.0	4.10
Median	1200000	12.0	6.18
Mean	1283238	11.79	5.78
Kuartil 3	1600000	12.00	7.50
Maksimum	8250000	120.00	14.77

Berdasarkan Tabel 2 diatas, bisa dilihat bahwa harga sewa memiliki rentang yang luas, mulai dari 350.000 Rupiah hingga 8.250.000 Rupiah dengan median berada di 1.200.000 Rupiah. Sebagian besar harga terpusat antara 800.000 Rupiah (Kuartil 1) dan 1.600.000 Rupiah (Kuartil 3). Harga sewa rata-rata yang lebih tinggi (1.283.238 Rupiah) dibandingkan median mengindikasikan adanya beberapa harga sewa indekos yang sangat tinggi yang mungkin berperan sebagai outlier menarik rata-rata ke atas. Ditinjau dari luas kamar juga cukup bervariasi dari 4.00 m² hingga 120.0 m² dengan nilai median sebesar 12.0 meter persegi. Sebagian besar kamar kos memiliki luas antara 9.00 m² (Kuartil 1) dan 12.00 m² (Kuartil 3), dengan rata-rata 11.79 m². Untuk variabel jarak, jarak dari lokasi kos ke kampus UPN berkisar antara 0.26 km hingga 14.77 km dengan median jarak 6.18 km. Sebagian besar indekos berada dalam jarak 4.10 km (Kuartil 1) hingga 7.50 km (Kuartil 3) dengan rata-rata jarak 5.78 km. Distribusi jarak ini menunjukkan bahwa mayoritas indekos terletak cukup dekat dengan kampus, meskipun ada beberapa kos yang terletak cukup jauh.

Dalam boxplot yang ditampilkan pada Gambar 2 tampak beberapa outlier. Pada data harga dan jarak, akan dihapus beberapa outlier dengan metode interquartile range (IQR). Terkait jarak yang berkaitan juga dengan lokasi, peneliti hanya akan menggunakan data indekos pada lokasi sekitaran Surabaya Timur. Hal ini berkaitan dengan patokan utama peneliti yakni UPN “Veteran” Jawa Timur yang terletak di Surabaya Timur. Dalam tahap pemodelan, variabel yang terkait lokasi (kategorik) tidak akan digunakan karena variabel ini akan diwakilkan nantinya dengan informasi *maps* ketika deployment. Beralih pada data luas, peneliti mencoba meninjau lebih lanjut data indekos yang diidentifikasi sebagai outlier. Berdasarkan pengecekan, didapatkan satu data indekos dengan outlier estrim (luas kamar 81). Pada data luas ini hanya satu data tersebut saja yang akan dihapus karena untuk data lain masih dalam cakupan luas yang wajar. Setelah

menyeleksi indekos dan menghapus outlier, dilakukan standarisasi pada ketiga variabel numerik ini supaya data memiliki skala yang sama.



Gambar 2. Boxplot Variabel Numerik

Berdasarkan Tabel 3 dibawah ini, terlihat bahwa fasilitas yang paling umum tersedia adalah Kasur (95.34%), WiFi (84.18%), fasilitas akses 24 jam (72.55%). Fasilitas AC tersedia di 59.68% indekos, sedangkan kamar mandi dalam ada di 53.17%. Sebaliknya, fasilitas yang paling jarang tersedia adalah listrik yang sudah termasuk dalam biaya sewa, dengan hanya 34.72% indekos yang menyediakannya. Melalui statistik deskriptif ini, peneliti memutuskan bahwa kolom terkait fasilitas WiFi, Kasur, dan Akses 24 Jam nantinya tidak akan digunakan dalam pemodelan karena mayoritas kos (>70%) menyediakan fasilitas tersebut. Hal ini dilakukan untuk menghindari bias dalam proses klusterisasi.

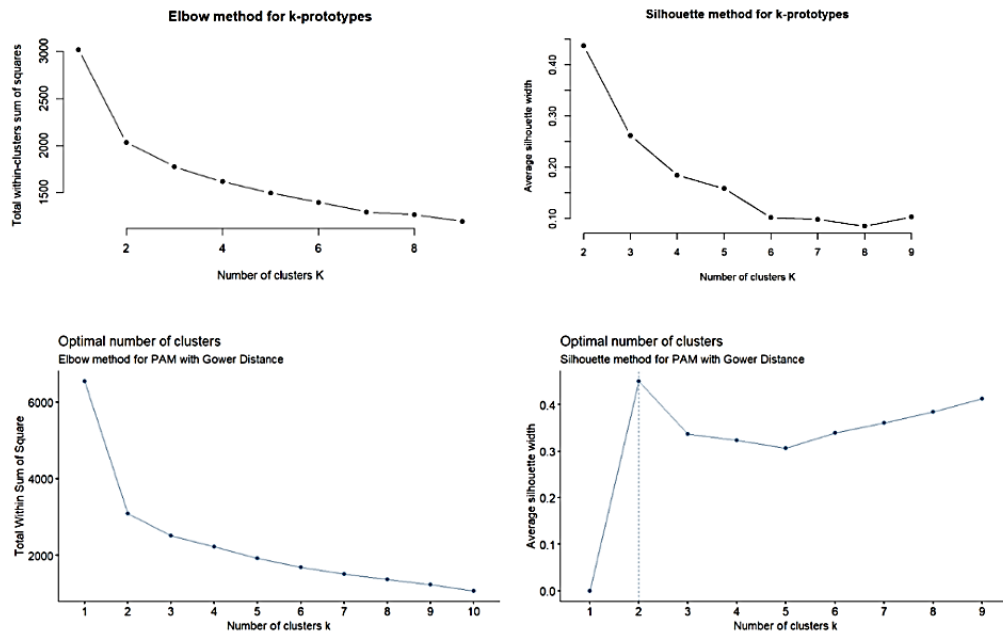
Tabel 3. Statistik Deskriptif Variabel Kategorik

Fasilitas	Persentase Ketersediaan (Yes)
K. Mandi Dalam	53.17 %
AC	59.68 %
WiFi	84.18 %
Kasur	95.34 %
Akses 24 Jam	72.55 %
Sudah Termasuk Listrik	34.72 %

Pemodelan dan Evaluasi

Setelah melakukan proses pembersihan dan eksplorasi, tahap berikutnya adalah pemodelan. Terdapat 473 *row* data dan 7 variabel (3 variabel numerik yang telah distandarisasi dan 4 variabel kategorik) yang akan digunakan dalam tahap ini. Adapun 7 variabel tersebut antara lain variabel Harga, Luas Kamar, Jarak ke UPN, Jenis, Listrik, K. Mandi Dalam, dan AC.

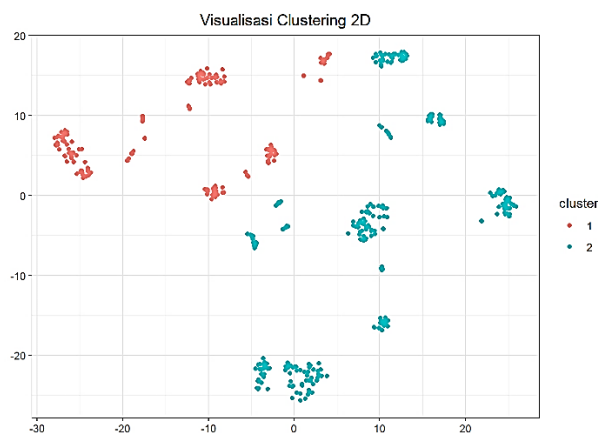
Ditinjau dari grafik Elbow pada Gambar 3, baik metode K-Prototype ataupun Partitioning Around Medoids (PAM) dengan Gower *Distance* memperlihatkan penurunan tajam dari K=1 ke K=2. Titik "siku" atau elbow pada kurva ini yang terlihat berada di sekitar K=2 mengindikasikan jumlah kluster yang optimum adalah 2 kluster. Begitupun dengan grafik *Silhouette*, nilai tertinggi masing-masing terdapat pada K=2. Namun, metode Partitioning Around Medoids (PAM) dengan Gower *Distance* memberikan nilai *Silhouette* yang sedikit lebih baik dibandingkan K-Prototype. Dalam hal ini, nilai *Silhouette* untuk metode Partitioning Around Medoids (PAM) dengan Gower *Distance* sebesar 0.450, sedangkan nilai *Silhouette* untuk metode K-Prototype ada di angka 0.436. Berdasarkan nilai ini, diputuskan bahwa metode Partitioning Around Medoids (PAM) dengan Gower *Distance* dan K = 2 akan digunakan untuk mengelompokkan data indekos.



Gambar 3. Grafik Elbow dan Silhouette

Interpretasi dan Visualisasi Kluster

Gambar 4 dibawah ini menunjukkan visualisasi yang dibangun dengan t-SNE (*t-distributed Stochastic Neighbor Embedding*). Dari Gambar 4 ini dapat diobservasi bahwa algoritma PAM telah berhasil membagi dua cluster yang berbeda dengan tidak ada yang saling tumpang tindih.



Gambar 4. Visualisasi t-SNE Klasterisasi Partitioning Around Medoids (PAM)

Dari proses klasterisasi, terdapat 184 (38.9%) kos tergolong dalam kluster 1 dan 289 (61.1%) masuk dalam kluster 2. Berdasarkan Gambar 5 dibawah ini kluster 1 dan kluster 2 memperlihatkan perbedaan yang signifikan dalam beberapa aspek. Pertama, dari segi harga, Kluster 1 memiliki median harga 750.000 Rupiah dan kuartil 3 sebesar 900.000 Rupiah, sementara kluster 2 memiliki median harga Rp 1.500.000 dengan kuartil 3 sebesar 1.800.000 Rupiah. Kedua, luas kamar juga menjadi pembeda yang jelas. Kluster 1 memiliki median luas kamar 9.00 m², sedangkan kluster 2 menawarkan ruang yang lebih luas dengan median 12.00 m². Perbedaan sekitar 3 m² ini cukup terasa dalam konteks hunian indekos, memberikan ruang gerak yang lebih leluasa bagi penghuni kluster 2. Ketiga, fasilitas yang disediakan juga menunjukkan perbedaan mencolok. Pada kluster 1 selain harganya jauh lebih murah, biaya ini juga sudah termasuk listrik.

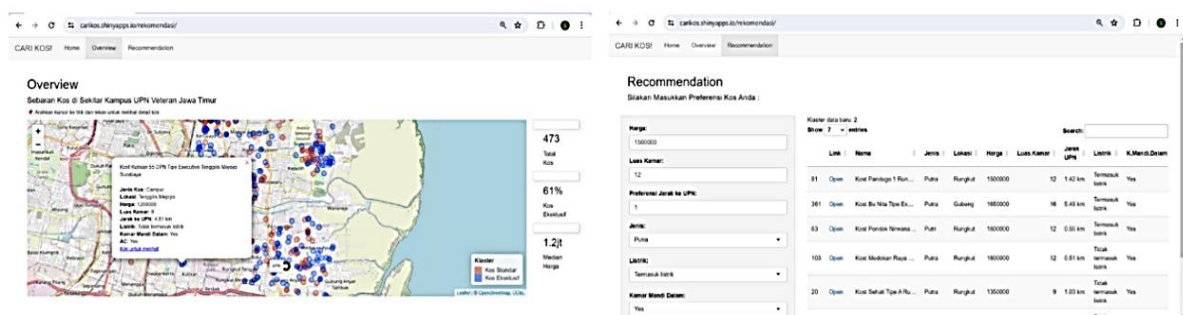
Berbeda dengan klaster 2 yang harganya jauh lebih mahal namun belum termasuk biaya listrik. Kemudian dari segi fasilitas kamar mandi dalam dan AC, mayoritas ada pada indekos klaster 2. Berdasarkan analisis ini, dapat disimpulkan bahwa klaster 1 mewakili segmen indekos dengan harga yang lebih terjangkau, ruang yang lebih *compact*, dan fasilitas yang lebih terbatas. Sedangkan klaster 2 merepresentasikan segmen indekos dengan eksklusif, ditandai dengan harga yang lebih mahal, kamar yang lebih luas dibanding klaster 1, serta dari segi fasilitas lebih lengkap. Beberapa kos di klaster 2 juga cenderung lebih dekat dengan kampus, hal ini dapat dilihat dari kuartil 1 pada klaster 2 yang jauh lebih kecil dibanding kuartil 1 pada klaster 1.

Klaster 1					Klaster 2				
pam2_results\$the_summary[[1]]					pam2_results\$the_summary[[2]]				
#>	Harga	Luas.Kamar	UPN	Jenis	#>	Harga	Luas.Kamar	UPN	Jenis
#> Min.	: 350000	Min. : 4.000	Min. :0.6006	Campur:15	#> Min.	: 500000	Min. : 4.00	Min. :0.2662	Campur: 75
#> 1st Qu.:	650000	1st Qu.: 9.000	1st Qu.:3.0668	Putra :70	#> 1st Qu.:	1200000	1st Qu.: 9.00	1st Qu.:1.7472	Putra : 68
#> Median :	750000	Median : 9.000	Median :5.0592	Putri :99	#> Median :	1500000	Median :12.00	Median :5.1228	Putri :146
#> Mean :	815766	Mean : 9.901	Mean :5.0036		#> Mean :	1497392	Mean :12.13	Mean :4.7906	
#> 3rd Qu.:	900000	3rd Qu.:12.000	3rd Qu.:7.2233		#> 3rd Qu.:	1800000	3rd Qu.:13.26	3rd Qu.:7.1270	
#> Max.	:2500000	Max. :30.000	Max. :9.9287		#> Max.	:2800000	Max. :32.00	Max. :9.8522	
#>		Listrik	K.Mandi.Dalam	AC	#>		Listrik	K.Mandi.Dalam	AC
#> Termasuk listrik	:135	No :167	No :162	Min. :1	#> Termasuk listrik	: 28	No : 69	No : 31	Min. :2
#> Tidak termasuk listrik:	49	Yes: 17	Yes: 22	1st Qu.:1	#> Tidak termasuk listrik:	261	Yes:220	Yes:258	1st Qu.:2
#>				Median :1	#>				Median :2
#>				Mean :1	#>				Mean :2
#>				3rd Qu.:1	#>				3rd Qu.:2
#>				Max. :1	#>				Max. :2

Gambar 5. Karakteristik Klaster 1 dan Klaster 2

Deployment dan User Acceptance Test (UAT)

Selanjutnya melakukan interpretasi klaster seperti yang telah diuraikan sebelumnya, langkah selanjutnya dalam menggunakan model klasterisasi untuk membangun sistem dengan R Shiny. Sistem ini nantinya akan memprediksi klaster dari data *input* preferensi pengguna dan memberikan rekomendasi indekos yang berada dalam klaster yang sama dengan *input* tersebut melalui perhitungan kedekatan Gower *Distance*. Gambar 6 berikut merupakan tampilan sistem yang telah dibangun.



Gambar 6. Tampilan Sistem

Untuk mengetahui kualitas sistem rekomendasi yang telah dibangun, dilakukan *User Acceptance Test* (UAT). Penyebaran kuisioner dilakukan secara acak pada mahasiswa Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur dengan total responden 104 mahasiswa. Ada 4 pertanyaan yang diberikan, P1: Rekomendasi yang dihasilkan sesuai dengan preferensi?, P2: Sistem responsif ketika diakses?, P3: Sistem sejenis ini mudah digunakan dalam menemukan kos yang sesuai preferensi?, P4: Pengalaman Anda secara keseluruhan sangat baik dalam menggunakan sistem ini?. Ada lima pilihan jawaban yang tersedia, yaitu Sangat Setuju (SS) dengan bobot nilai 5, Setuju (S) dengan nilai 4, Cukup Setuju (CS) dengan nilai 3, Kurang Setuju (KS) dengan nilai 2, dan Tidak Setuju (TS) dengan nilai 1.

Tabel 4. Rekapitulasi Jawaban Responden

	SS	S	CS	KS	TS	% SS	% S	% CS	% KS	% TS
P1	30	60	14	0	0	28.85	57.69	13.46	0	0
P2	59	38	7	0	0	56.73	36.54	6.73	0	0
P3	51	39	14	0	0	49.04	37.50	13.46	0	0
P4	54	43	7	0	0	51.92	41.35	6.73	0	0
Total	194	180	42	0	0					
Total × Bobot	970	360	126	0	0					

Jumlah skor observasi adalah jumlah dari skor masing-masing pertanyaan hasil observasi dikalikan dengan bobot skor menurut skala *likert* ditunjukkan pada hasil Total x Bobot. Skor maksimal dapat dihitung dengan mengalikan nilai tertinggi pada skala *likert* dengan jumlah pertanyaan, sehingga diperoleh $5 \times 4 = 20$. Jumlah skor yang diharapkan merupakan hasil perkalian antara skor maksimal dengan jumlah responden, yaitu $20 \times 104 = 2080$. Sehingga, persentase kebaikan sistem dapat dihitung dengan:

$$\text{Persentase Kebaikan Sistem} = \frac{\sum \text{Total x Bobot}}{\text{Skor yang diharapkan}} \times 100\% = \frac{1456}{2080} \times 100\% = 70\%$$

KESIMPULAN DAN SARAN

Kesimpulan

Pengembangan sistem rekomendasi melalui pendekatan analisis kluster terbukti berhasil dalam penelitian ini. Hal ini bisa dilihat dari hasil klusterisasi dan nilai *User Acceptance Test* (UAT). Nilai *Silhouette* untuk metode Partitioning Around Medoids (PAM) dengan Gower Distance adalah sebesar 0.450 yang berarti model mampu membentuk kluster-kluster yang cukup kompak dan terpisah dengan baik satu sama lain. Proses klusterisasi ini membagi data ke dalam dua kluster. Terdapat 184 (38.9%) kos tergolong dalam kluster 1 (kos standar) dan 289 (61.1%) masuk dalam kluster 2 (kos eksklusif). Model ini kemudian digunakan untuk mengembangkan sistem rekomendasi berbasis preferensi menggunakan kedekatan jarak Gower. Hasil pengujian sistem rekomendasi melalui *User Acceptance Test* (UAT) menunjukkan nilai 70%, mengindikasikan bahwa secara keseluruhan sistem dapat dikatakan baik dan dapat diterima oleh pengguna.

Saran

Saran yang dapat disampaikan peneliti untuk mengembangkan penelitian ini dimasa yang akan datang adalah memperluas cakupan objek penelitian dan mencoba membandingkan metode klusterisasi lainnya serta konsep jarak tertentu dalam membangun sistem rekomendasi yang berbasis preferensi pengguna.

REFERENSI

- Ashari, I. F., Dwi Nugroho, E., Baraku, R., Novri Yanda, I., & Liwardana, R. (2023). Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta. *Journal of Applied Informatics and Computing*, 7(1), 89–97. <https://doi.org/10.30871/jaic.v7i1.4947>
- Astiyana Putra, N. P., & Harianto, S. (2022). Konstruksi Sosial Mahasiswa Urban Di Kota Surabaya. *Jurnal Analisa Sosiologi*, 11(2), 306–325. <https://doi.org/10.20961/jas.v11i2.57277>
- Bartasari, L., & Arnellis, A. (2023). Preferensi Mahasiswa UNP Terhadap Tempat Kos di kec. Padang Utara pada Masa Pandemi (Covid-19) dengan Menggunakan Analisis Konjoin. *Journal of Mathematics UNP*, 8(1), 1. <https://doi.org/10.24036/unpjomath.v8i1.12730>

- Dewa, F. A., & Jatipaningrum, M. T. (2019). Segmentasi E-Commerce dengan cluster K-Means dan Fuzzy C-means (Studi kasus: media sosial di Indonesia yang diunduh di Play Store). *Jurnal Statistika Industri dan Komputasi*, 4(1), 53–67. <https://doi.org/10.34151/statistika.v4i01.1054>
- Fadilah, Z. R., & Wijayanto, A. W. (2023). Perbandingan Metode Klasterisasi Data Bertipe Campuran: One-Hot-Encoding, Gower Distance, dan K-Prototype Berdasarkan Akurasi (Studi Kasus: Chronic Kidney Disease Dataset). *Journal of Applied Informatics and Computing*, 7(1), 57–67. <https://doi.org/10.30871/jaic.v7i1.5857>
- Haryati, A. E., & Sugiyarto. (2021). Clustering with Principal Component Analysis and Fuzzy Subtractive Clustering Using Membership Function Exponential and Hamming Distance. *IOP Conference Series: Materials Science and Engineering*, 1077(1), 012019. <https://doi.org/10.1088/1757-899x/1077/1/012019>
- Kusumantara, P. M., Kirana, F. H., & Fahmi, A. (2020). Implementasi Metode Profile Matching Untuk Pendukung Keputusan Pemilihan Indeks Mahasiswa. *SCAN - Jurnal Teknologi Informasi dan Komunikasi*, 15(2), 34–39. <https://doi.org/10.33005/scan.v15i2.2042>
- Muhaimin, A. (2018). *Deteksi Anomali Pada Pemakaian Air Pelanggan Pdam Surya Sembada Kota Surabaya Menggunakan Algoritma Kohonen Self Organizing Maps (Som) Dan Local Outlier Factor (Lof)* [Institut Teknologi Sepuluh Nopember.]. <https://repository.its.ac.id/57048/>
- Nooraeni, R., Suprijadi, J., & Zulhanif. (2019). K-Prototype untuk Pengelompokan Data Campuran. *Jurnal Statistika Teori dan Aplikasi: Biomedics, Industry & Business And Social Statistics*, 13(1), 9–16.
- Paruntu, J. G., Hatidja, D., & Langi, Y. A. . (2021). Keputusan Mahasiswa dalam Memilih Indeks dengan Analisis Faktor. *Jurnal Ilmiah Sains*, 21(2), 119. <https://doi.org/10.35799/jis.v21i2.28868>
- Prasath, V. B. S., Alfeilat, H. A. A., Hassanat, A. B. A., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., & Salman, H. S. E. (2017). *Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier -- A Review*. 1–50. <https://doi.org/10.1089/big.2018.0175>
- Prasetya, D. A., Nguyen, P. T., Faizullin, R., Iswanto, I., & Armay, E. F. (2020). Resolving the shortest path problem using the haversine algorithm. *Journal of Critical Reviews*, 7(1), 62–64. <https://doi.org/10.22159/jcr.07.01.11>
- Sipayung, E. M., Fiarni, C. F., & Sutopo, S. (2021). Sistem Rekomendasi Tempat Kost di Sekitar Kampus ITHB Menggunakan Metode Analytical Hierarchy Process (AHP). *Jurnal Nasional Teknologi dan Sistem Informasi*, 7(2), 52–60. <https://doi.org/10.25077/teknosi.v7i2.2021.52-60>
- Wibawa, E. G., & Dewa, P. K. (2024). Boarding House Property Market Trends and Investor Preferences in Boarding House Development: A Comparative Study with Web Scraping. *Jurnal Rekayasa Industri (JRI)*, 6(1), 85–99. <https://doi.org/10.37631/jri.v6i1.1379>
- Wijoyo, S., & Maimunah, E. (2019). Faktor-faktor Pertimbangan Mahasiswa UNILA dalam Pemilihan Rumah Indeks dikelurahan Kampung Baru dan Gedung Meneng Bandar Lampung. *Jurnal Ekonomi Pembangunan*, 8(1), 45–55. <https://doi.org/10.23960/jep.v8i1.35>