

**ANALISIS KLASTER *PARTITIONING AROUND MEDOIDS*
(PAM) DENGAN *GOWER DISTANCE* UNTUK
REKOMENDASI INDEKOS
(Studi Kasus: Indekos di Wilayah Sekitar Kampus UPNVJT)**

SKRIPSI

**Diajukan untuk memenuhi salah satu syarat kelulusan
di Program Studi Sains Data**



Disusun Oleh:

SAHAT RENALDI. S

20083010026

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2024**

LEMBAR PENGESAHAN

**ANALISIS KLASTER *PARTITIONING AROUND MEDOIDS* (PAM)
DENGAN *GOWER DISTANCE* UNTUK REKOMENDASI INDEKOS**

(Studi Kasus: Indekos di Wilayah Sekitar Kampus UPNVJT)

SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains Data
pada : Senin, 15 Juli 2024

**Program Studi S-1 Sains Data
Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jawa Timur
Surabaya**

Oleh :

SAHAT RENALDI. S

NPM. 20083010026

Disetujui oleh Tim Penguji Skripsi :

Penguji 1

Trimono, S.Si., M.Si.
NIP. 199509082022031003
Pembimbing 1

Penguji 2

Tresna Maulana Fahrudin, S.ST., M.T.
NIP. 199305012022031007
Pembimbing 2

Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU.
NIP. 198012052005011002

Amri Muhaimin, S.Stat., M.Stat., M.S.
NPT. 21119950723270

Mengetahui,

Fakultas Ilmu Komputer
Dekan,

Prof. Dr. Ir. Novirina Hendrasarie, M.T.
NIP. 196811261994032001

Program Studi Sains Data
Fakultas Ilmu Komputer
Koordinator,

Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU.
NIP. 198012052005011002

Surabaya, Juli, 2024

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : Sahat Renaldi. S
NPM : 20083010026
Program Studi : Sains Data

Menyatakan bahwa judul Skripsi / Tugas Akhir sebagai berikut:

**“ANALISIS KLASSTER *PARTITIONING AROUND MEDOIDS* (PAM)
DENGAN *GOWER DISTANCE* UNTUK REKOMENDASI INDEKOS
(Studi Kasus: Indekos di Wilayah Sekitar Kampus UPNVJT)”**

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk/ *software*/ hasil karya yang saya beli dari orang lain.

Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka, dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata dikemudian hari pernyataan ini terbukti tidak benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari.

Surabaya, 22 Juli 2024

Hormat Saya,



Sahat Renaldi. S

NPM. 20083010026

ABSTRAK

ANALISIS KLASTER *PARTITIONING AROUND MEDOIDS* (PAM) DENGAN *GOWER DISTANCE* UNTUK REKOMENDASI INDEKOS (STUDI KASUS: INDEKOS DI SEKITAR KAMPUS UPNVJT)

Nama Mahasiswa / NPM : Sahat Renaldi. S / 20083010026
Program Studi : Sains Data, FASILKOM, UPN Veteran Jatim
Dosen Pembimbing 1 : Dr.Eng.Ir.Dwi Arman Prasetya, S.T., M.T., IPU.
Dosen Pembimbing 2 : Amri Muhaimin, S.Stat., M.Stat., M.S.

Abstrak

Surabaya merupakan kota metropolitan paling besar kedua setelah Jakarta dan memegang peran penting dalam banyak aspek kehidupan nasional, termasuk dari aspek pendidikan. Berdasarkan data dari BPS Jatim (2022), terdapat 70 perguruan tinggi di Kota Surabaya. Kondisi ini mendorong migrasi pendidikan. Setiap tahun, ribuan mahasiswa baru dari berbagai latar belakang bermigrasi ke kota besar seperti Surabaya untuk melanjutkan pendidikan. Fenomena tersebut berdampak pada peningkatan kebutuhan akan properti indekos. Selain meningkatnya kebutuhan terhadap akomodasi rumah kos (indekos), beragamnya pilihan dan fasilitas yang tersedia di berbagai lokasi sering kali menjadi tantangan bagi mahasiswa dalam memilih indekos yang memenuhi ekspektasi dan pilihan individual. Situasi ini memerlukan pendekatan analitis yang dapat mendukung individu dalam menemukan indekos yang sesuai dengan preferensi mereka. Penelitian ini bertujuan untuk merancang sistem rekomendasi indekos dengan memanfaatkan metode analisis kluster *K-Prototype* dan *Partitioning Around Medoids (PAM)* yang dikombinasikan 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. Metode klusterisasi PAM dengan *Gower Distance* menunjukkan kinerja lebih unggul dengan perolehan indeks *Silhouette* sebesar 0.45 dan berhasil membagi indekos ke dalam 2 kluster.

Klaster 1 mewakili segmen indeks standar dengan harga yang relatif lebih terjangkau, ruang yang relatif lebih sempit, dan fasilitas yang lebih terbatas. Sedangkan klaster 2 merepresentasikan segmen indeks eksklusif, ditandai dengan harga yang relatif lebih mahal, area kamar yang relatif lebih luas, dan fasilitas yang lebih komprehensif/lengkap dibanding klaster 1. Sistem rekomendasi dikembangkan berdasarkan konsep *Gower Distance*. Hasil pengujian sistem melalui *User Acceptance Test* (UAT) didapatkan skor sebesar 86.73% yang menunjukkan bahwa sistem ini secara keseluruhan dapat dianggap baik dan diterima oleh pengguna.

Kata kunci: Indeks, *K-Prototype*, *Partitioning Around Medoids* (PAM), Sistem Rekomendasi, *Gower Distance*.

ABSTRACT

PARTITIONING AROUND MEDOIDS (PAM) CLUSTERING ANALYSIS USING GOWER DISTANCE FOR BOARDING HOUSE RECOMMENDATIONS

(CASE STUDY: BOARDING HOUSES AROUND UPNVJT CAMPUS)

Student Name / NPM : Sahat Renaldi. S / 20083010026
Study Program : Sains Data, FASILKOM, UPN Veteran Jatim
Advisor 1 : Dr.Eng.Ir.Dwi Arman Prasetya, S.T., M.T., IPU.
Advisor 2 : Amri Muhaimin, S.Stat., M.Stat., M.S.

Abstract

Surabaya is the second largest metropolis after Jakarta and holds an important role in many aspects of national life, including the educational sector. According to data from BPS East Java (2022), there are 70 universities in the city of Surabaya. This condition encourages educational migration. Every year, thousands of new students from diverse backgrounds migrate to large cities like Surabaya to continue their education. This phenomenon has resulted in a heightened demand for boarding houses properties. In addition to the increasing demand for boarding house accommodations, the diverse range of options and facilities available in various locations often presents challenges for students in selecting the most suitable boarding house that aligns with their individual needs and preferences. This situation requires an analytical approach that can assist individuals in finding a boarding house while taking into account their preferences. The objective of this study is to create a recommendation system by employing K-Prototype and Partitioning Around Medoids (PAM) techniques, utilizing Gower Distance. The object of this research is boarding houses located in the East Surabaya region, particularly in the vicinity of the UPN "Veteran" Jawa Timur campus, which serves as the primary reference point for this study. The PAM method with Gower Distance performed better, achieving a Silhouette index of

0.45, and successfully klastered boarding houses into two categories. Cluster 1 represents the segment of standard boarding houses with more affordable prices, smaller room spaces, and more limited facilities. In contrast, Cluster 2 represents the exclusive indekos segment, characterized by higher prices, more spacious rooms, and more complete facilities. The recommendation system is based on Gower Distance similarity. User Acceptance Test (UAT) results showed a score of 86.73%, indicating that the system is generally well-received and acceptable to users.

Keywords: Boarding House, K-Prototype, Partitioning Around Medoids (PAM), Recommendation System, Gower Distance.

KATA PENGANTAR

Segala Puji dan Syukur penulis hanturkan kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya yang telah memungkinkan penulis menyelesaikan penelitian ini. Buku ini disusun sebagai perjalanan akhir dalam pembelajaran sebagai mahasiswa di Program Studi S1-Sains Data, Fakultas Ilmu Komputer, UPN “Veteran” Jawa Timur. Melalui kesempatan ini, penulis ingin memberikan ucapan terima kasih kepada kepada semua pihak yang telah berkontribusi, terutama kepada:

1. Kedua orangtua dan segenap saudara yang senantiasa mendoakan penulis dan selalu mendukung serta memberikan semangat.
2. Prof. Dr. Ir. Akhmad Fauzi, M.MT, yang menjabat sebagai Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur, atas dukungan dan dedikasi yang telah diberikan.
3. Prof. Dr. Ir. *Novirina Hendrasarie*, M.T, selaku Dekan Fakultas Ilmu Komputer.
4. Bapak Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU, yang merupakan Koordinator Program Studi Sains Data sekaligus sebagai Dosen Pembimbing I, atas motivasi, arahan dan dorongan kepada penulis.
5. Bapak Amri Muhaimin, S.Stat., M.Stat., M.S selaku Dosen Pembimbing II yang telah memberikan banyak pencerahan dan masukan yang berkelanjutan selama proses penelitian skripsi ini.
6. Sivitas Akademika di Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Semoga hasil penelitian tugas akhir ini dapat memperluas pengetahuan bagi pihak yang membacanya.

Surabaya, 3 Mei 2024

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	ii
SURAT PERNYATAAN	iii
ABSTRAK	iv
ABSTRACT	vi
KATA PENGANTAR.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
DAFTAR LAMPIRAN.....	xiii
BAB I. PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian	6
BAB II. TINJAUAN PUSTAKA.....	7
2.1. Dasar Teori	7
2.1.1. Indekos	7
2.1.2. Profil Kota Surabaya	7
2.1.3. <i>Web Scraping</i>	8
2.1.4. Analisis Deskriptif	9
2.1.5. <i>Machine Learning</i>	11
2.1.6. Analisis Klaster	12
2.1.7. Analisis Klaster Berbasis Partisi	12
2.1.8. Ukuran Jarak	13
2.1.9. <i>K-Prototype</i>	23
2.1.10. <i>Partitioning Around Medoids</i> (PAM)	27
2.1.11. Evaluasi Klaster	31
2.1.12. <i>User Acceptance Test</i> (UAT).....	33

2.2.	Penelitian Terdahulu	35
BAB III. METODOLOGI PENELITIAN		39
3.1.	Variabel Penelitian dan Sumber Data	39
3.2.	Langkah Analisis.....	40
3.3.	Diagram Alir Penelitian	44
3.4.	Jadwal Penelitian.....	45
BAB IV. HASIL DAN PEMBAHASAN		46
4.1	Pengumpulan dan Pembersihan Data.....	46
4.2	Analisis Deskriptif dan Pra-Pemodelan	54
4.3	Pemodelan dan Evaluasi	57
4.4	Interpretasi dan Visualisasi Kluster.....	59
4.5	<i>Deployment</i>	61
4.6	<i>User Acceptance Test (UAT)</i>	68
BAB V. PENUTUP.....		71
5.1	Kesimpulan	71
5.2	Saran.....	71
DAFTAR PUSTAKA		72
LAMPIRAN.....		80
BIODATA PENULIS.....		83

DAFTAR GAMBAR

Gambar 1. 1 Sistem berbasis Filter pada Salah Satu <i>Platform</i> Indekos	2
Gambar 2. 1 Ilustrasi <i>Web Scraping</i>	9
Gambar 2. 2 Ilustrasi Proses <i>Machine Learning</i>	11
Gambar 2. 3 Metode Analisis Kluster	12
Gambar 2. 4 Parameter ISO 9126-3	34
Gambar 3. 1 Tampilan <i>robots.txt</i> Mamikos.....	39
Gambar 3. 2 Diagram Alir Penelitian.....	44
Gambar 4. 1 <i>Pop-Up</i> di <i>Multi Pages</i> Mamikos	47
Gambar 4. 2 <i>Pop-Up</i> di <i>Single Page</i> Mamikos	48
Gambar 4. 3 Cuplikan Data Hasil <i>Scraping</i>	51
Gambar 4. 4 Variabel Penelitian.....	54
Gambar 4. 5 Boxplot Variabel Numerik	55
Gambar 4. 6 Data Pemodelan.....	57
Gambar 4. 7 Grafik Elbow dan <i>Silhouette</i>	58
Gambar 4. 8 Distribusi Kluster Indekos Berdasarkan Lokasi	61
Gambar 4. 9 Alur Kerja Sistem Rekomendasi	62
Gambar 4. 10 Tampilan Menu <i>Overview</i>	64
Gambar 4. 11 Tampilan Menu <i>Recommendation</i>	64
Gambar 4. 12 Pengujian 1	65
Gambar 4. 13 Pengujian 2	66
Gambar 4. 14 Pengujian 3	67

DAFTAR TABEL

Tabel 2. 1	Cara Kerja Algoritma Klasterisasi Berbasis Partisi	13
Tabel 2. 2	Data Perhitungan Jarak Euclidean.....	14
Tabel 2. 3	Data Perhitungan Jarak <i>Simple Matching</i>	16
Tabel 2. 4	Data Perhitungan Jarak Gower.....	20
Tabel 2. 5	Data Perhitungan <i>K-Prototype</i>	25
Tabel 2. 6	Hasil Perhitungan <i>K-Prototype</i> pada Iterasi 1	26
Tabel 2. 7	Titik <i>Centroid</i> Baru untuk Iterasi 2	26
Tabel 2. 8	Hasil Perhitungan <i>K-Prototype</i> pada Iterasi 2.....	27
Tabel 2. 9	Data Perhitungan <i>Cost</i> PAM.....	29
Tabel 2. 10	Hasil Perhitungan Jarak ke <i>Medoid</i>	29
Tabel 2. 11	Hasil Perhitungan Jarak ke <i>Medoid</i> Baru.....	30
Tabel 2. 12	Pilihan Jawaban Menurut Skala <i>Likert</i>	34
Tabel 2. 13	Konversi Hasil Penilaian	34
Tabel 2. 14	Penelitian Terdahulu	35
Tabel 3. 1	Data Indekos.....	40
Tabel 3. 2	Jadwal Penelitian.....	45
Tabel 4. 1	Statistik Deskriptif Variabel Numerik Keseluruhan Data.....	54
Tabel 4. 2	Persentase Ketersediaan Fasilitas.....	56
Tabel 4. 3	Statistik Deskriptif Variabel Numerik Klaster 1	59
Tabel 4. 4	Statistik Deskriptif Variabel Numerik Klaster 2	59
Tabel 4. 5	Persentase Ketersediaan Fasilitas	60
Tabel 4. 6	Persentase Kos Berdasarkan Jenis dan Klaster	60
Tabel 4. 7	Pertanyaan <i>User Acceptance Test</i>	69
Tabel 4. 8	Rekapitulasi Jawaban Responden	69

DAFTAR LAMPIRAN

Lampiran 1. Hasil Uji Plagiasi	80
Lampiran 2. Dokumentasi <i>Source Code</i>	81
Lampiran 3. Data Penelitian	82