

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Pemerintah Kota Surabaya senantiasa mengupayakan kesejahteraan masyarakat. Hal ini terwujud dalam program “Dandan Omah” yang selaras dengan upaya pembangunan berkelanjutan. Program ini merupakan program rehabilitasi rumah tidak layak huni (*Rutilahu*) bagi warga Kota Surabaya yang kurang mampu. Menurut Dinas Perumahan Rakyat dan Kawasan Permukiman serta Pertanahan (DPRKPP) Kota Surabaya banyak keluarga di Surabaya yang berhadapan dengan rumah di bawah standar dan memerlukan perbaikan besar. Ada sekitar 2,51 juta unit rumah tidak layak huni. Dari total tersebut, sebanyak 2,18 juta rumah berisiko menjadi tidak layak huni, sementara 330.000 rumah sudah dianggap benar-benar tidak layak huni. Skema Rehabilitasi *Rutilahu* yang digagas pemerintah bertujuan untuk mengatasi permasalahan semacam ini. Dengan adanya program bantuan rehabilitasi rumah dari pemerintah ini, masyarakat antusias menyambut dengan berbondong-bondong mengurus berkas pengajuan bantuan. Hal ini terlihat dari data Dinas Perumahan Rakyat dan Kawasan Permukiman serta Pertanahan (DPRKPP) Kota Surabaya yang menunjukkan adanya kenaikan pengajuan bantuan tiap tahunnya mulai dari tahun 2020 [1].

Berdasarkan data dari Dinas Perumahan Rakyat dan Kawasan Permukiman serta Pertanahan (DPRKPP) Kota Surabaya, jumlah pengajuan rehabilitasi *Rutilahu* semakin meningkat setiap tahunnya. Tahun 2017, terdapat 1.629 orang yang melakukan pengajuan. Tahun 2018, terdapat 1.648 orang yang melakukan pengajuan. Namun, di tahun 2019 dan 2020 pengajuan bantuan ini terlihat menurun. Hal ini disebabkan adanya pembatasan kuota rehabilitasi oleh pemerintah Kota Surabaya. Pembatasan ini dilakukan untuk mengalokasikan dana rehabilitasi *Rutilahu* ke dalam upaya percepatan penurunan Covid-19. Pada Tahun 2019 terdapat 1.090 pengajuan dan tahun 2020 terdapat 463 pengajuan. Kemudian, di tahun 2021 hingga 2022, jumlah pengajuan mulai meningkat. Terdapat 842 pengajuan di tahun 2021 dan 950 pengajuan di tahun 2022. Pada tahun 2023, jumlah

pengajuan mencapai puncaknya, tercatat terdapat 2.700 orang melakukan pengajuan bantuan ini [2].

Skema program bantuan ini dibuat oleh pemerintah dalam mengupayakan kesejahteraan masyarakatnya. Namun, proses penentuan penerima bantuan rehabilitasi *Rutilahu* banyak sekali tantangan. Selama ini, proses penentuan penerima bantuan ini sering kali tidak tepat pada sasaran. Hal ini disebabkan oleh penilaian yang masih bersifat subjektif dalam proses seleksi penerimaan bantuan. Subjektivitas tersebut muncul dikarenakan kriteria penilaian yang belum terstandarisasi secara objektif. Sehingga penerima bantuan sering kali ditentukan berdasarkan pertimbangan yang tidak sepenuhnya berbasis data atau fakta lapangan. Hal ini mengakibatkan, terlewatnya banyak pihak yang seharusnya layak menerima bantuan, sementara yang tidak begitu membutuhkan bisa saja terpilih sebagai penerima.

Oleh sebab itu, diperlukan sistem otomatisasi untuk mengelompokkan masyarakat yang berhak menerima bantuan rehabilitasi *Rutilahu* tersebut. Perancangan sistem ini merupakan proses merumuskan langkah-langkah teknis dan metodologis yang diperlukan untuk membangun sebuah sistem berbasis data yang mampu berjalan secara otomatis [3]. Pengelompokan dilakukan dengan mengelolah data penduduk yang melakukan proses pengajuan bantuan di kelurahan. Pengelompokan ini dilakukan dengan mempertimbangkan variabel sosial dan ekonomi yang sudah ditetapkan. Selain itu, pengelompokan ini juga bertujuan untuk menentukan calon penerima bantuan mana saja yang diprioritaskan perbaikannya. Sehingga rumah yang benar-benar tidak layak huni dapat disegerakan proses rehabilitasinya.

Permasalahan rumah tidak layak huni merupakan isu sosial yang tidak hanya terbatas pada konteks lokal, tetapi juga menjadi perhatian global dalam upaya pencapaian *Sustainable Development Goals* (SDGs) yang menekankan pada kota dan permukiman yang inklusif, aman, dan berkelanjutan [4]. *Dataset Rutilahu* yang digunakan dalam penelitian ini mencerminkan upaya pemerintah dalam menanggulangi ketimpangan sosial melalui program perbaikan rumah. Secara teknis, *dataset* ini memiliki kombinasi atribut numerik dan kategorikal ordinal, menjadikannya tantangan tersendiri dalam proses klusterisasi. Kompleksitas ini

memberikan ruang eksplorasi yang relevan untuk pengujian algoritma optimasi dalam konteks *clustering* data campuran.

*Clustering* merupakan metode dalam *data mining* yang bertujuan untuk mengelompokkan data ke dalam kelompok berdasarkan kesamaan karakteristiknya. Metode ini digunakan untuk menemukan pola atau struktur tersembunyi pada data yang tidak memiliki label. Salah satu keunggulan utama pengelompokan data adalah kemampuannya mengenali pola secara otomatis tanpa memerlukan informasi sebelumnya [5]. Salah satu metode *clustering* yang dirancang untuk menangani data campuran adalah algoritma *K-Prototypes*. Metode ini menggabungkan pendekatan algoritma *K-Means* yang biasa digunakan untuk menangani data numerikal dan *K-Modes* yang biasa digunakan untuk menangani data kategorikal [6]. Dengan menggunakan *K-Prototypes*, pengelompokan data menjadi lebih relevan dan akurat ketika menghadapi *dataset* yang kompleks. *K-Prototypes* telah terbukti memberikan hasil yang lebih baik dibandingkan metode *clustering* lainnya dalam konteks data campuran. Namun, performa algoritma *K-Prototypes* sangat bergantung pada inisialisasi awal pusat klaster. Jika inisialisasi kurang optimal, hasil *clustering* bisa tidak maksimal [7]. Oleh karena itu, diperlukan metode optimasi untuk meningkatkan akurasi dan efektivitas algoritma ini.

Optimasi pada *clustering* bertujuan untuk meningkatkan kualitas hasil pengelompokan, seperti mengurangi kesalahan dan meningkatkan akurasi. Dalam penelitian ini, tiga algoritma optimasi dipertimbangkan untuk mengatasi kelemahan *K-Prototypes*, di antaranya adalah *Particle Swarm Optimization* (PSO), *Genetic Algorithm* (GA), dan *Flower Pollination Algorithm* (FPA). PSO adalah algoritma optimasi yang terinspirasi oleh perilaku sosial burung dalam mencari makanan. Dalam konteks *clustering*, PSO dapat digunakan untuk mengoptimalkan posisi pusat klaster dalam algoritma *clustering*, sehingga menghasilkan pengelompokan yang lebih akurat [8]. Sementara itu, GA merupakan metode optimasi yang didasarkan pada prinsip seleksi alam, pada prosesnya individu-individu dalam populasi akan berevolusi untuk menghasilkan solusi yang lebih baik. Dalam konteks *clustering*, GA dapat digunakan untuk mengoptimalkan parameter algoritma *clustering*, seperti jumlah klaster dan posisi pusat klaster, sehingga dapat

meningkatkan kualitas pengelompokan [9]. Di sisi lain, FPA mengambil inspirasi dari proses penyerbukan bunga, di mana pola penyerbukan yang efisien dapat digunakan untuk menemukan solusi optimal dalam pengelompokan data [10]. Metode ini menawarkan pendekatan yang unik dan efektif untuk meningkatkan performa *clustering* dengan memanfaatkan mekanisme alami. Dengan memanfaatkan ketiga metode optimasi ini, penelitian ini bertujuan untuk mengeksplorasi dan membandingkan efektivitas PSO, FPA, dan GA dalam meningkatkan hasil *clustering*.

Penelitian ini berfokus pada tiga algoritma optimasi utama, yaitu PSO, GA, dan FPA dengan mempertimbangkan rekam jejak penggunaannya dalam konteks pengelompokan data campuran. Terdapat studi-studi terdahulu yang secara eksplisit menerapkan GA dan FPA untuk mengoptimalkan pusat kluster pada algoritma *K-Prototypes*. Hal ini menunjukkan potensi keduanya dalam menangani data dengan atribut numerik dan kategorikal secara bersamaan. Di sisi lain, PSO menunjukkan tingkat adopsi yang cukup tinggi dalam berbagai penelitian terkait segmentasi data sosial, termasuk pengelompokan calon penerima bantuan, meskipun belum banyak diintegrasikan secara langsung dengan *K-Prototypes*. Oleh sebab itu, eksplorasi lebih lanjut terhadap ketiga algoritma ini dalam skenario yang seragam dan berbasis data sosial campuran menjadi relevan untuk dilakukan, guna memperoleh pemahaman yang lebih mendalam mengenai karakteristik dan efektivitas masing-masing pendekatan dalam meningkatkan kualitas hasil klusterisasi.

Penelitian yang dilakukan oleh Wijaya et al. (2024) mengoptimasi pemilihan pusat kluster awal pada algoritma *K-Medoids* menggunakan *Particle Swarm Optimization* (PSO) untuk segmentasi pelanggan. *K-Medoids* dipilih karena keunggulannya dalam mengurangi sensitivitas terhadap *outlier* dibandingkan *K-Means*, namun masih memiliki kelemahan pada proses inisialisasi pusat kluster yang acak dan rentan terhadap *local optima*. Untuk mengatasi hal tersebut, PSO digunakan guna menentukan pusat kluster awal yang lebih optimal. Hasil evaluasi menggunakan *Davies-Bouldin Index* menunjukkan bahwa metode PSO+*K-Medoids* memberikan hasil klusterisasi yang lebih baik dengan nilai DBI sebesar 0,088 pada  $k = 2$ , dibandingkan *K-Medoids* standar yang memperoleh DBI 0,283 pada  $k = 3$ .

Hal ini membuktikan bahwa integrasi PSO mampu meningkatkan kualitas hasil klasterisasi secara signifikan dalam segmentasi pelanggan [11].

Penelitian yang dilakukan oleh Suwirmayati et al. (2020) melakukan optimasi pada pusat klaster *K-Prototypes* menggunakan *Genetic Algorithm* (GA) dalam menangani data *German Credit* yang bertipe campuran. Selain itu, penelitian ini juga membandingkan kinerja algoritma *K-Prototypes* konvensional, *K-Prototypes* dengan optimasi GA, dan *K-Means*. Berdasarkan hasil penelitian yang dilakukan, menunjukkan bahwa *K-Prototypes* dengan optimasi GA adalah metode yang lebih baik dibandingkan *K-Prototypes* konvensional dan *K-Means*. Hal ini terlihat pada total *cost* atau total jarak. Metode *K-Prototypes* dengan optimasi GA menghasilkan nilai *cost* terendah, sebesar 2622. Sedangkan *K-Prototypes* konvensional memiliki nilai *cost* sebesar 2706 dan *K-Means* memiliki nilai *cost* sebesar 9579 [7].

Penelitian optimasi *K-Prototypes* juga dilakukan oleh Mulyadewi et al. (2024) yang memanfaatkan metode *Flower Pollination Algorithm* (FPA) untuk mengoptimalkan proses klasterisasi pada *dataset* penderita diabetes melitus tipe 2 di Kota Gorontalo. Algoritma *K-Prototypes* dipilih karena kemampuannya menangani data campuran yang terdiri dari atribut numerik dan kategorikal, sementara FPA digunakan untuk menentukan pusat klaster terbaik berdasarkan nilai *fitness*. Hasil penelitian menunjukkan bahwa kombinasi *K-Prototypes* dan FPA menghasilkan 3 klaster terbaik dengan nilai *fitness* sebesar -0,501 yang menunjukkan kualitas pengelompokan yang cukup optimal. Secara lebih rinci, klaster 1 terdiri dari 9 kelurahan, klaster 2 dari 11 kelurahan, dan klaster 3 dari 30 kelurahan [12].

Meskipun beberapa penelitian sebelumnya menunjukkan peningkatan kualitas *clustering* berdasarkan nilai *Davies-Bouldin Index* (DBI) atau *fitness*, perlu dikaji lebih lanjut alasan pemilihan metrik-metrik tersebut dalam konteks data campuran seperti data *Rutilahu*. DBI memang umum digunakan karena mengukur seberapa baik masing-masing klaster terpisah dan kompak [13]. Namun, dalam penelitian ini juga dipertimbangkan metrik evaluasi tambahan berupa *Silhouette Score*, yang tidak hanya mengukur kohesi dan separasi antar klaster, tetapi juga memberikan nilai evaluasi per data *point* serta memungkinkan visualisasi distribusi

klaster [14]. Hal ini mendukung analisis yang lebih komprehensif dan interpretatif, terutama dalam konteks data sosial yang membutuhkan pemeriksaan granular terhadap hasil klasterisasi.

Dari beberapa penelitian terdahulu, menunjukkan efektivitas algoritma optimasi seperti PSO, GA, dan FPA dalam meningkatkan kualitas klasterisasi. Namun, penerapannya masih banyak difokuskan pada skenario dengan kompleksitas data yang terbatas. Dalam konteks sistem berskala besar, diperlukan pendekatan yang tidak hanya menghasilkan klasterisasi yang akurat, tetapi juga efisien secara komputasi dan adaptif terhadap keterbatasan sumber daya. Penelitian ini mengusulkan strategi optimasi yang lebih efisien dan fleksibel terhadap variasi struktur data, guna mendukung penerapan yang lebih luas dalam lingkungan operasional yang kompleks.

Penelitian ini memiliki pendekatan holistik dalam optimasi pusat klaster *K-Prototypes* menggunakan tiga algoritma metaheuristik secara bersamaan, yaitu PSO, GA, dan FPA, yang belum pernah dibandingkan secara langsung dalam konteks data campuran penerima bantuan sosial. Selain itu, implementasi sistem klasterisasi berbasis antarmuka pengguna (GUI) dengan Streamlit yang memungkinkan pengambilan keputusan secara otomatis, cepat, dan berbasis data oleh pihak kelurahan atau pemangku kebijakan. Pendekatan yang menggabungkan penggunaan beberapa algoritma optimasi serta penerapan sistem interaktif ini ditujukan untuk memperkuat analisis dan pemanfaatan data dalam konteks penyaluran bantuan sosial.

Ketidaktepatan dalam proses seleksi penerima bantuan rehabilitasi rumah tidak layak huni (*Rutilahu*) di Kota Surabaya berpotensi menimbulkan ketidakadilan sosial dan inefisiensi distribusi bantuan. Peningkatan jumlah pengajuan dari masyarakat setiap tahunnya menunjukkan perlunya sistem pengelompokan yang mampu menangani data dalam skala besar secara objektif dan terstandarisasi. Ketergantungan pada penilaian manual yang bersifat subjektif menjadi hambatan serius dalam mewujudkan distribusi bantuan yang tepat sasaran. Dalam kondisi ini, pendekatan berbasis data melalui metode *clustering* yang dioptimasi menggunakan algoritma metaheuristik seperti *Particle Swarm Optimization* (PSO), *Genetic Algorithm* (GA), dan *Flower Pollination Algorithm*

(FPA) menjadi sangat relevan untuk diterapkan. Penggunaan pendekatan ini tidak hanya meningkatkan akurasi pengelompokan, tetapi juga mendukung upaya pemerintah dalam membangun sistem seleksi yang transparan dan akuntabel. Hal tersebut selaras dengan agenda pembangunan berkelanjutan (*Sustainable Development Goals/SDGs*) pada aspek permukiman layak dan pengurangan ketimpangan.

## 1.2. Rumusan Masalah

Berdasarkan uraian latar belakang diatas, adapun rumusan masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut.

- a. Bagaimana cara membuat model *K-Prototypes clustering* untuk data penerimaan bantuan rehabilitasi *Rutilahu* di Kota Surabaya?
- b. Bagaimana cara mengoptimalkan model *K-Prototypes clustering* agar menghasilkan pengelompokan yang lebih akurat?
- c. Bagaimana cara mengevaluasi kinerja model *K-Prototypes clustering* yang telah dibangun?
- d. Bagaimana cara mengimplementasikan model *K-Prototypes clustering* terbaik ke dalam bentuk antarmuka pengguna (GUI)?

## 1.3. Batasan Masalah

Untuk menjaga fokus serta validitas penelitian, terdapat beberapa batasan masalah yang ditetapkan, diantaranya:

- a. Penelitian ini menggunakan data penerima bantuan rehabilitasi rumah tidak layak huni (*Rutilahu*) di Kota Surabaya.
- b. Penelitian ini menggunakan metode *K-Prototypes* untuk pengelompokan data campuran (numerik dan kategorik).
- c. Pada penelitian ini, pemilihan parameter untuk PSO, GA, dan FPA dilakukan berdasarkan serangkaian uji coba parameter dengan metode eksperimen. Nilai parameter diuji dengan menaikkan dan menurunkan nilainya dalam rentang tertentu untuk mengamati pengaruhnya terhadap hasil *clustering*.

- d. Metode evaluasi kinerja pengelompokan yang menggunakan *Davies-Bouldin Index* (DBI) untuk mengukur kepadatan dan separasi antar klaster, *Silhouette Score* untuk menilai seberapa baik pemisahan antar klaster, serta waktu komputasi untuk mengevaluasi efisiensi algoritma.

#### 1.4. Tujuan Penelitian

Berdasarkan rumusan masalah dan batasan penelitian yang sudah ditetapkan, adapun tujuan dari penelitian ini adalah untuk mengoptimalkan model *K-Prototypes clustering* dalam mengelompokkan penerima bantuan rehabilitasi *Rutilahu* di Kota Surabaya, sehingga dapat meningkatkan akurasi dalam penentuan kelompok penerima yang tepat sasaran.

#### 1.5. Manfaat Penelitian

Dari rumusan masalah dan tujuan yang ingin dicapai dari penelitian, adapun manfaat dari penelitian ini adalah sebagai berikut.

- a. Untuk membantu pemerintah Kota Surabaya dalam mengelompokkan penerima bantuan rehabilitasi *Rutilahu* dengan lebih tepat dan efisien.
- b. Untuk meningkatkan akurasi dan efektivitas *clustering* yang dapat digunakan sebagai acuan dalam pemilihan metode optimasi yang paling tepat.
- c. Untuk memberikan wawasan mengenai metode optimasi yang paling efektif untuk diterapkan pada *K-Prototypes clustering* dalam pengelompokan penerimaan bantuan rehabilitasi *Rutilahu*.
- d. Untuk memudahkan pengguna non-teknis dalam mengetahui hasil pengelompokan penerimaan bantuan rehabilitasi *Rutilahu* pada bentuk antarmuka (GUI).
- e. Untuk membantu pemerintah dalam menyalurkan bantuan dengan lebih transparan dan efisien, serta mengurangi potensi kesalahan atau ketidakadilan dalam distribusi bantuan.