

BAB II

LANDASAN TEORI

2.1. Kelayakan Hunian

Kelayakan hunian merujuk pada sejauh mana suatu tempat tinggal memenuhi standar yang diperlukan untuk kesehatan, kenyamanan, dan keselamatan penghuninya [5]. Faktor-faktor yang menentukan kelayakan hunian meliputi ketersediaan infrastruktur dasar seperti air bersih, sanitasi, sistem drainase, aksesibilitas jalan, serta kondisi fisik bangunan yang harus memenuhi syarat keamanan [6]. Selain itu, faktor sosial, ekonomi, dan lingkungan juga berperan penting, di mana kawasan pemukiman yang ideal harus memiliki akses ke fasilitas pendidikan, kesehatan, dan pasar yang memadai, serta tidak berada di lokasi rawan bencana atau polusi.

Dalam penelitian ini, penerapan algoritma *K-Means* digunakan untuk mengelompokkan kawasan pemukiman berdasarkan kelayakan hunian yang diukur dari berbagai variabel tersebut [7]. Dengan pendekatan ini, data tentang faktor-faktor yang memengaruhi kelayakan hunian seperti infrastruktur, fasilitas umum, dan kondisi lingkungan dapat dikelompokkan secara efisien, sehingga menghasilkan pemetaan kawasan yang lebih terstruktur. Hasil pengelompokan ini diharapkan dapat memberikan informasi yang berguna bagi Dinas Perumahan dan Kawasan Pemukiman dalam mengambil keputusan terkait pengembangan kawasan yang lebih layak huni.

2.2. Dinas Perumahan dan Kawasan Permukiman

Dinas Perumahan dan Kawasan Permukiman (DPKP) memiliki peran yang sangat penting dalam pembangunan dan pengelolaan kawasan perumahan di suatu daerah. Tugas utamanya adalah merancang, mengawasi, serta memastikan keberlanjutan pengembangan perumahan yang layak huni, baik untuk masyarakat berpenghasilan rendah (MBR) maupun untuk kalangan lainnya [1]. Selain itu, DPKP juga berperan dalam penyusunan kebijakan, perencanaan tata ruang, serta pengembangan infrastruktur yang mendukung kualitas hidup masyarakat, seperti akses air bersih, sanitasi, dan jalan. Dalam menjalankan tugasnya, DPKP bekerja sama dengan berbagai lembaga lain untuk memastikan bahwa pembangunan perumahan tidak hanya memenuhi standar teknis, tetapi juga mendukung kelestarian lingkungan dan kesejahteraan sosial.

Penelitian ini berfokus pada penerapan algoritma *K-Means* dalam pengelompokan kawasan permukiman berdasarkan kelayakan hunian. Algoritma ini digunakan untuk menganalisis data yang terkait dengan berbagai aspek kelayakan hunian, seperti kualitas infrastruktur, aksesibilitas dan kondisi lingkungan. Dengan menggunakan metode *Clustering*, kawasan permukiman dapat dikelompokkan secara otomatis ke dalam kategori-kategori tertentu, memudahkan pihak Dinas Perumahan dan Kawasan Permukiman dalam mengambil keputusan terkait perencanaan dan pengembangan kawasan perumahan. Hal ini juga memberikan wawasan yang lebih mendalam mengenai wilayah yang membutuhkan perhatian lebih dalam hal peningkatan fasilitas dan pengelolaan sumber daya untuk menciptakan hunian yang lebih layak.

2.3. *Knowledge Discovery in Database* (KDD)

Knowledge Discovery in Database (KDD) adalah proses yang digunakan untuk mengekstraksi pengetahuan yang berguna dari data besar yang tersimpan dalam database [8]. Proses ini melibatkan serangkaian langkah yang dimulai dari pengumpulan data, pra-pemrosesan untuk membersihkan dan mempersiapkan data, pemilihan fitur yang relevan, hingga penerapan algoritma *Data mining* untuk mengidentifikasi pola atau hubungan tersembunyi. KDD bertujuan untuk menemukan informasi yang tidak hanya dapat meningkatkan pemahaman terhadap data tetapi juga memberikan wawasan yang berguna bagi pengambilan Keputusan. KDD menjadi penting dalam konteks analisis data yang kompleks karena membantu menggali nilai tersembunyi yang mungkin tidak terdeteksi dengan cara konvensional.

Dalam penelitian ini, pendekatan KDD digunakan untuk mengelompokkan kawasan pemukiman berdasarkan kelayakan hunian. Langkah pertama dalam proses KDD adalah pengumpulan data yang mencakup berbagai variabel terkait kondisi pemukiman, seperti infrastruktur, kualitas lingkungan, dan aksesibilitas [9]. Setelah itu, data akan diproses dan dipersiapkan agar algoritma *K-Means* dapat diterapkan untuk mengidentifikasi zona pemukiman dengan kualitas hunian yang serupa. Dengan menggunakan KDD, penelitian ini dapat membantu menghasilkan rekomendasi yang lebih akurat dalam merencanakan dan meningkatkan kualitas hunian, serta memberikan wawasan yang lebih jelas bagi pihak yang bertanggung jawab dalam pengelolaan kawasan pemukiman.

2.1.1. *Data mining*

Data mining adalah proses untuk menemukan pola, hubungan, atau informasi yang berguna dari sejumlah besar data yang tersimpan dalam database [10]. Dengan menggunakan teknik statistik, algoritma, dan model matematis, *Data mining* membantu dalam menggali wawasan dari data yang tidak terstruktur atau tersembunyi. Teknik-teknik dalam *Data mining* termasuk klasifikasi, regresi, *Clustering*, dan asosiasi, yang semuanya memiliki tujuan untuk mengidentifikasi pola yang dapat digunakan untuk prediksi atau pengambilan keputusan. Dalam dunia bisnis, *Data mining* sering digunakan untuk memahami perilaku konsumen, memprediksi tren pasar, serta meningkatkan efisiensi operasional. Seiring berkembangnya teknologi, kemampuan untuk menganalisis data dalam jumlah besar semakin penting dalam mengambil keputusan yang lebih tepat dan berbasis data.

Dalam penelitian ini, penerapan algoritma *K-Means* sebagai salah satu metode *Clustering* dalam *Data mining* digunakan untuk mengelompokkan kawasan pemukiman berdasarkan kelayakan hunian [11]. Dengan menggunakan *K-Means*, data pemukiman yang meliputi berbagai atribut, seperti infrastruktur, lingkungan, dan fasilitas, dikelompokkan ke dalam beberapa kategori untuk mempermudah analisis kelayakan. Proses ini memungkinkan identifikasi kawasan yang memenuhi kriteria kelayakan hunian dan yang tidak, sehingga dapat menjadi dasar bagi pengambilan keputusan terkait pengembangan kawasan pemukiman. Penggunaan *Data mining* melalui *K-Means* memberikan kemudahan dalam mengelompokkan kawasan dengan cara yang efisien dan objektif, memberikan gambaran yang lebih

jasas bagi pihak terkait dalam merencanakan pengembangan perumahan yang lebih baik.

2.1.2. Database dan Data Processing

Database adalah sistem yang dirancang untuk mengelola, menyimpan, dan mengorganisir data secara efisien agar dapat diakses dan dikelola dengan mudah [12]. Database biasanya terdiri dari tabel-tabel yang berisi data terstruktur dan dapat dihubungkan satu sama lain melalui hubungan tertentu. Sistem manajemen basis data (DBMS) seperti MySQL, PostgreSQL, dan MongoDB memungkinkan pengguna untuk melakukan operasi seperti penyimpanan, pengambilan, pembaruan, dan penghapusan data dengan cepat dan aman. Selain itu, database juga mendukung berbagai metode pengolahan data, seperti query SQL, indeks pencarian, serta optimasi penyimpanan untuk meningkatkan kinerja dan skalabilitas sistem.

Dalam proses pengolahan data, database memainkan peran penting dalam berbagai penelitian yang membutuhkan analisis data dalam jumlah besar. Data yang tersimpan dalam database perlu diproses dengan metode tertentu agar menghasilkan informasi yang bermakna [13]. Data processing mencakup tahapan seperti pembersihan data (data cleaning), transformasi data, analisis statistik, serta penerapan algoritma untuk pengelompokan atau prediksi. Misalnya, dalam penelitian yang menggunakan algoritma *K-Means*, database digunakan untuk menyimpan data kawasan pemukiman, yang kemudian diproses untuk mengidentifikasi kelompok berdasarkan kelayakan hunian. Dengan teknik pengolahan data yang tepat, penelitian dapat menghasilkan wawasan yang lebih akurat dan bermanfaat bagi pengambilan keputusan.

2.1.3. *Visualization*

Visualization adalah teknik penyajian data dalam bentuk grafis untuk mempermudah pemahaman pola, tren, dan hubungan dalam data [14]. Dengan menggunakan berbagai bentuk visual seperti diagram, grafik, peta, dan infografik, *Visualization* membantu dalam menyederhanakan data yang kompleks sehingga lebih mudah dianalisis. Dalam analisis data, *Visualization* memainkan peran penting dalam mendukung pengambilan keputusan, karena memungkinkan interpretasi cepat dan akurat dari informasi yang tersedia. Selain itu, perkembangan teknologi telah memungkinkan pembuatan visualisasi interaktif yang lebih dinamis, sehingga pengguna dapat mengeksplorasi data secara lebih mendalam dan fleksibel.

Dalam penerapan algoritma *K-Means*, *Visualization* digunakan untuk menampilkan hasil *Clustering* agar lebih mudah dipahami dan dianalisis. Salah satu metode yang sering digunakan adalah scatter plot dengan warna berbeda untuk setiap cluster, yang membantu dalam melihat pola pengelompokan secara visual [15]. Selain itu, teknik seperti heatmap dan elbow method juga dapat digunakan untuk menentukan jumlah cluster yang optimal. *Visualization* juga berperan dalam memvalidasi hasil *Clustering* dengan membandingkan distribusi data sebelum dan sesudah pengelompokan, sehingga dapat memastikan bahwa metode yang diterapkan memberikan hasil yang sesuai dengan tujuan analisis.

2.1.4. Statistik

Statistik adalah cabang ilmu yang berfokus pada pengumpulan, analisis, interpretasi, dan penyajian data untuk memahami pola serta mengambil keputusan berdasarkan informasi yang tersedia [16]. Dalam statistik, terdapat dua jenis utama, yaitu statistik deskriptif dan statistik inferensial. Statistik deskriptif digunakan

untuk menggambarkan dan merangkum data melalui ukuran seperti mean, median, dan standar deviasi, sementara statistik inferensial digunakan untuk membuat prediksi atau kesimpulan dari sampel terhadap populasi yang lebih besar. Berbagai metode statistik memungkinkan pengolahan data menjadi lebih bermakna, sehingga dapat digunakan dalam berbagai bidang, seperti ekonomi, kesehatan, social dan teknologi.

Dalam penelitian, statistik berperan penting dalam validasi data dan pengambilan keputusan berbasis bukti. Misalnya, dalam analisis data menggunakan algoritma seperti *K-Means*, statistik membantu dalam menentukan jumlah *Cluster* optimal dan mengevaluasi hasil pengelompokan [17]. Statistik juga digunakan dalam analisis regresi, uji hipotesis serta pengukuran akurasi model untuk memastikan hasil yang diperoleh dapat diandalkan. Dengan menerapkan teknik statistik yang tepat, penelitian dapat menghasilkan kesimpulan yang lebih akurat dan mendukung pengambilan keputusan yang lebih baik dalam berbagai bidang.

2.1.5. Pattern recognition

Pattern recognition adalah cabang kecerdasan buatan yang berfokus pada identifikasi pola dalam data dan pengambilan keputusan berdasarkan pola tersebut. Metode ini digunakan untuk mengklasifikasikan objek ke dalam kategori tertentu dengan memanfaatkan teknik statistik, machine learning, dan deep learning [13]. Proses *Pattern recognition* melibatkan beberapa tahap, seperti ekstraksi fitur, pemrosesan data, dan klasifikasi. Implementasi *Pattern recognition* dapat ditemukan dalam berbagai bidang, termasuk pengenalan wajah, analisis teks, deteksi anomaly dan pengelompokan data dalam sistem informasi.

Dalam penelitian, *Pattern recognition* banyak digunakan untuk menganalisis data dan menemukan pola tersembunyi yang dapat membantu pengambilan Keputusan [18]. Salah satu metode yang sering diterapkan adalah algoritma *K-Means*, yang mengelompokkan data berdasarkan kemiripan karakteristik. Misalnya, dalam analisis pemukiman, *K-Means* dapat membantu dalam mengelompokkan kawasan berdasarkan kelayakan hunian dengan mempertimbangkan faktor seperti infrastruktur, aksesibilitas, dan kondisi lingkungan. Dengan adanya *Pattern recognition*, penelitian dapat menghasilkan informasi yang lebih akurat dan bermanfaat dalam berbagai bidang, termasuk perencanaan kota, Kesehatan dan bisnis.

2.4. Metode *K-Means*

Metode *K-Means* adalah salah satu algoritma dalam *Data mining* yang digunakan untuk melakukan *Clustering* atau pengelompokan data berdasarkan kesamaan karakteristik. Algoritma ini bekerja dengan menentukan sejumlah pusat cluster (centroid) secara acak, kemudian menghitung jarak setiap data ke centroid terdekat menggunakan Euclidean Distance [19]. Data yang memiliki jarak paling dekat akan dikelompokkan dalam cluster yang sama. Proses ini berulang hingga posisi centroid stabil dan tidak berubah lagi. Keunggulan *K-Means* terletak pada efisiensinya dalam mengelompokkan data dalam jumlah besar dengan cepat, meskipun algoritma ini memiliki kelemahan seperti ketergantungan pada pemilihan jumlah cluster yang tepat dan kemungkinan hasil yang berbeda jika pemilihan centroid awal berubah.

Dalam penerapannya, metode *K-Means* banyak digunakan untuk mengelompokkan data dalam berbagai bidang, seperti analisis wilayah rawan

bencana, pengelompokan pelanggan dalam bisnis, hingga klasifikasi kawasan pemukiman. Dengan *K-Means*, data pemukiman dapat dikelompokkan berdasarkan parameter tertentu, seperti kualitas infrastruktur, aksesibilitas, atau tingkat kepadatan penduduk, sehingga dapat membantu pengambilan keputusan dalam perencanaan tata kota [20]. Algoritma ini memberikan hasil yang lebih objektif dibandingkan pendekatan manual karena berbasis pada analisis pola dari data yang tersedia. Hal ini memungkinkan pemerintah atau pihak terkait untuk menetapkan kebijakan pembangunan yang lebih tepat sasaran, seperti prioritas peningkatan infrastruktur di wilayah yang kurang layak huni.

2.5. Model *Clustering*

Model *Clustering* adalah salah satu metode dalam *Data mining* yang digunakan untuk mengelompokkan data berdasarkan kemiripan atau karakteristik tertentu. Berbeda dengan metode klasifikasi yang menggunakan label yang sudah ditentukan sebelumnya, *Clustering* bekerja tanpa label (*unsupervised learning*) dan menemukan pola alami dalam data [20]. Algoritma dalam *Clustering*, seperti *K-Means*, DBSCAN, dan Hierarchical *Clustering*, bertujuan untuk memisahkan data ke dalam kelompok yang homogen di dalamnya tetapi heterogen dibandingkan dengan kelompok lain. Proses ini melibatkan pemilihan metrik jarak, jumlah *Cluster* yang optimal, serta validasi hasil *Clustering* menggunakan metode seperti Silhouette Score atau *Davies-Bouldin Index*.

Penerapan model *Clustering* sangat berguna dalam berbagai bidang, termasuk analisis kawasan pemukiman, segmentasi pelanggan, deteksi anomali, hingga pengelompokan tingkat kesejahteraan masyarakat [21]. Dengan menggunakan algoritma *K-Means*. Misalnya, kawasan pemukiman dapat dikelompokkan

berdasarkan kelayakan hunian dengan mempertimbangkan faktor seperti infrastruktur, aksesibilitas, dan kondisi lingkungan. Hasil *Clustering* ini dapat membantu dalam pengambilan keputusan seperti perencanaan pembangunan, alokasi sumber daya serta peningkatan kualitas hunian di berbagai wilayah.

2.6. Alat Bantu Program Aplikasi RapidMiner

RapidMiner adalah salah satu perangkat lunak yang digunakan untuk analisis data dan *Data mining* tanpa memerlukan keterampilan pemrograman yang mendalam. Aplikasi ini menyediakan antarmuka berbasis drag-and-drop yang memungkinkan pengguna dengan mudah membangun model analisis, mulai dari pra-pemrosesan data, eksplorasi pola, hingga penerapan algoritma pembelajaran mesin [22]. RapidMiner mendukung berbagai metode *Data mining*, seperti klasifikasi, *Clustering*, regresi dan asosiasi. Serta kompatibel dengan berbagai format data. Dengan fitur yang kaya dan integrasi yang fleksibel, RapidMiner menjadi pilihan populer dalam berbagai bidang penelitian dan industri untuk mengolah data secara efisien dan akurat.

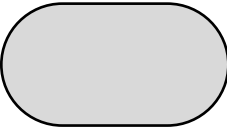
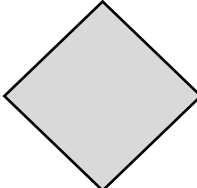
Dalam penelitian yang menggunakan algoritma *K-Means*, RapidMiner dapat dimanfaatkan untuk melakukan segmentasi atau pengelompokan data berdasarkan kesamaan karakteristik. Prosesnya mencakup pemuatan dataset, normalisasi data, pemilihan jumlah *Cluster* optimal, serta visualisasi hasil *Clustering* [23]. Dengan bantuan RapidMiner, analisis menjadi lebih sistematis karena aplikasi ini menyediakan berbagai metrik evaluasi yang membantu dalam menentukan kualitas *Cluster* yang terbentuk. Selain itu, kemampuannya dalam menangani data dalam jumlah besar menjadikannya alat yang efektif untuk penelitian yang memerlukan

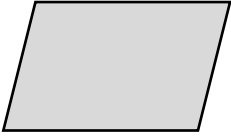
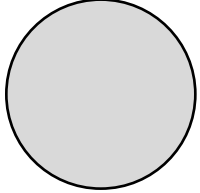
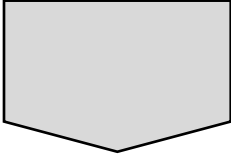
pengelompokan berbasis data dalam berbagai bidang, termasuk perumahan, ekonomi, kesehatan, dan bisnis.

2.7. Flowchart

Flowchart adalah diagram yang digunakan untuk memvisualisasikan alur kerja atau proses secara sistematis dengan simbol-simbol standar. Dalam penelitian ini, flowchart digunakan untuk menggambarkan tahapan penerapan algoritma K-Means dalam pengelompokan kawasan permukiman berdasarkan kelayakan hunian.

Tabel 2. 1. Flowchart

	Simbol Flowchart	Nama Simbol	Arti Simbol
1		Terminator	Simbol ini menunjukkan awal atau akhir dari suatu proses dalam flowchart. Contoh: Terminator digunakan untuk menandai titik mulai atau selesai dalam suatu alur proses.
2		Process	Simbol ini menggambarkan langkah atau tindakan yang dilakukan dalam suatu proses. Contoh: Process digunakan untuk menunjukkan operasi atau aktivitas yang terjadi dalam sistem.
3		Document	Simbol ini digunakan untuk menunjukkan pengeluaran atau hasil dalam bentuk dokumen atau laporan. Contoh: Document digunakan untuk menggambarkan pembuatan atau pencetakan dokumen dalam alur proses.
4		Decision	Simbol ini digunakan untuk menunjukkan titik keputusan atau kondisi dalam alur, yang menghasilkan cabang keputusan. Contoh: Decision digunakan untuk menentukan jalur alur

			berdasarkan kondisi tertentu, seperti "Ya" atau "Tidak."
5		Data	Simbol ini menunjukkan input atau output data dalam sistem. Contoh: Data digunakan untuk menggambarkan pengambilan data dari pengguna atau menampilkan hasil dari proses.
6		On-Page Reference/Connect or	Simbol ini digunakan untuk menghubungkan bagian-bagian flowchart yang ada di halaman yang sama, memudahkan pemahaman alur yang lebih kompleks. Contoh: On-Page Reference digunakan untuk menghubungkan langkah-langkah dalam flowchart yang tersebar di halaman yang sama.
7		Off-Page Reference/Odd- Page Connector	Simbol ini digunakan untuk menghubungkan bagian flowchart yang terletak di halaman berbeda atau dalam alur yang lebih besar. Contoh: Off-Page Reference digunakan untuk merujuk ke proses atau langkah yang ada di halaman lain dalam flowchart yang lebih besar.
8		Flow	Simbol ini menunjukkan arah aliran proses atau data dalam flowchart. Contoh: Flow digunakan untuk mengarahkan pembaca mengikuti langkah-langkah proses sesuai urutannya.