

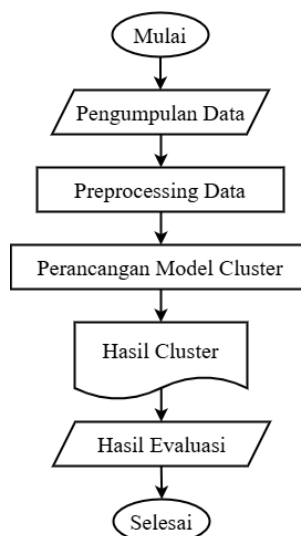
BAB III METODOLOGI PENELITIAN

3.1. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode *Data mining* dengan algoritma *K-Means* untuk melakukan pengelompokan kawasan pemukiman berdasarkan kelayakan hunian [24]. Tahapan penelitian dimulai dengan pengumpulan data terkait faktor-faktor yang mempengaruhi kelayakan hunian, seperti aksesibilitas, ketersediaan fasilitas umum, kondisi infrastruktur, dan aspek lingkungan. Data yang diperoleh kemudian melalui proses pra-pemrosesan, termasuk pembersihan data, normalisasi, dan transformasi agar siap untuk dianalisis. Setelah itu, algoritma *K-Means* diterapkan dengan menentukan jumlah *Cluster* yang optimal menggunakan metode evaluasi seperti Elbow Method atau Silhouette Score [25].

3.2. Kerangka Kerja Penelitian

Pada penelitian ini terdapat kerangka kerja penelitian yang dirancang dengan menggunakan Flowchart yaitu sebagai berikut.



Gambar 3. 1. Flowchart

1. Mulai, Tahapan awal yang menandai dimulainya proses penelitian. Pada tahap ini, peneliti menentukan tujuan dan ruang lingkup penelitian serta menyiapkan segala kebutuhan teknis, termasuk perangkat lunak dan data yang akan dianalisis.
2. Pengumpulan Data, Pada tahap ini, data yang dibutuhkan untuk penelitian dikumpulkan dari sumber yang telah ditentukan, misalnya melalui survei, dokumen instansi, atau data sekunder yang relevan. Data yang dikumpulkan mencakup atribut-atribut penting seperti jenis kelamin, aksesibilitas infrastruktur, dan kondisi sosial ekonomi.
3. Preprocessing Data, Tahapan ini bertujuan untuk menyiapkan data agar layak digunakan dalam proses analisis. Dalam konteks penelitian ini, karena data sudah cukup rapi, proses preprocessing hanya melibatkan transformasi data dari bentuk kategorikal menjadi numerikal untuk memudahkan pengolahan oleh algoritma K-Means.
4. Perancangan Model Cluster, Pada tahap ini dilakukan proses perancangan dan penerapan algoritma clustering, dalam hal ini menggunakan metode K-Means. Model diuji dengan dua variasi jumlah *Cluster*, yaitu $K = 3$ dan $K = 5$, untuk mengetahui struktur pengelompokan yang paling optimal.
5. Hasil Cluster, Tahap ini menyajikan hasil dari proses clustering, yaitu pengelompokan data ke dalam *Cluster-Cluster* tertentu berdasarkan kemiripan karakteristik. Hasil ditampilkan dalam bentuk tabel yang mencantumkan setiap individu beserta *Cluster* yang dihasilkannya untuk masing-masing nilai K .
6. Hasil Evaluasi, Pada tahapan ini dilakukan evaluasi terhadap model clustering yang telah dibuat, menggunakan metrik seperti Average Within Centroid

Pengamatan ini bertujuan untuk memperoleh data faktual yang mendukung analisis dalam Penelitian [26].

2. Wawancara dilakukan dengan pihak Dinas Perumahan dan Kawasan Permukiman untuk mendapatkan informasi mendalam mengenai kebijakan, program, serta kriteria kelayakan hunian yang diterapkan [27].

3.7. Teknik Analisis Data

Teknik analisis data dalam penelitian ini menggunakan aplikasi RapidMiner untuk menentukan nilai *Cluster* dengan algoritma *K-Means*. Data yang telah diproses dimasukkan ke dalam RapidMiner, di mana jumlah *Cluster* optimal ditentukan menggunakan metode evaluasi seperti Elbow Method [28]. Hasil *Clustering* kemudian dianalisis untuk mengidentifikasi pola dan karakteristik masing-masing kelompok permukiman berdasarkan kelayakan huniannya [29].

3.8. Perhitungan Data

3.8.1. Pengumpulan Data

Tahapan pengumpulan data merupakan langkah awal yang dilakukan dalam penelitian ini dengan tujuan untuk mengumpulkan informasi yang akan dijadikan dasar dalam proses analisis. Data yang digunakan dalam penelitian ini berjumlah sebanyak 10 data, yang masing-masing merepresentasikan objek permukiman berdasarkan tiga variabel utama, yaitu infrastruktur dasar, aksesibilitas, dan kondisi sosial ekonomi. Seluruh data tersebut kemudian digunakan dalam proses clustering guna mengelompokkan kawasan permukiman berdasarkan kemiripan karakteristiknya.

Tabel 3. 2. Data Sampel Penelitian

Nama	Jenis Kelamin	Aksesibilitas Infrastruktur	Kondisi Sosial Ekonomi	NIK	Alamat	Tipe Bangunan	Jalan
Adrian Putra	Laki-Laki	Sangat Baik	Sedang	1213076592183740	Padang Matinggi	42	Beton
Bayu Kurniawan	Laki-Laki	Baik	Tinggi	1218927365012640	Padang Matinggi	36	Beton
Candra Wijaya	Laki-Laki	Sangat Baik	Tinggi	1211763928471520	Padang Matinggi	42	Beton
Denny Prasetyo	Laki-Laki	Baik	Rendah	1213647859206170	Padang Matinggi	42	Beton
Evan Saputra	Laki-Laki	Kurang Baik	Rendah	1215782039461580	Padang Matinggi	42	Beton
Farhan Maulana	Laki-Laki	Sangat Baik	Sedang	1219056217483200	Padang Matinggi	42	Beton
Gilang Ramadhan	Laki-Laki	Sangat Baik	Tinggi	1214398271059860	Padang Matinggi	42	Beton
Hendra Setiawan	Laki-Laki	Sangat Baik	Tinggi	1216283741905270	Padang Matinggi	42	Beton
Iqbal Permana	Laki-Laki	Sangat Baik	Sedang	1213195628471090	Padang Matinggi	42	Beton
Junaidi Akbar	Laki-Laki	Baik	Tinggi	1217946352184060	Padang Matinggi	36	Beton

Pada tabel diatas merupakan data sampel yang akan digunakan pada penelitian ini. Data diatas merupakan data yang diperoleh dari dinas Perumahan dan Kawasan Pemukiman.

3.8.2. Preprocessing Data

Pada tahapan preprocessing data pada penelitian ini merupakan tahapan yang dilakukan untuk merubah data sampel di konversi pada bilangan numerikal. Untuk keterangan data konversinya yaitu sebagai berikut.

Tabel 3. 3. Data Keterangan Atribut

Atribut	Partisi	Nilai
Akseibilitas Infrastruktur	Sangat Baik	5
	Baik	3
	Kurang Baik	2
Kondisi Sosial Ekonomi	Tinggi	50
	Sedang	30
	Rendah	20

Pada tabel diatas merupakan data keterangan yang digunakan sebagai keterangan untuk data yang dikonversi pada penelitian ini. Untuk data yang sudah dikonversi yang itu sebagai berikut.

1. Akseibilitas Infrastruktur

Atribut ini mengacu pada kemudahan akses ke fasilitas umum seperti jalan, air bersih, listrik, sekolah, puskesmas, dan sarana lainnya. Tingkat akseibilitas menunjukkan seberapa mudah penghuni rumah menjalani kehidupan sehari-hari secara fungsional.

Partisi	Nilai	Penjelasan
Sangat Baik	5	Rumah berada di lokasi strategis, dekat jalan utama, mudah dijangkau kendaraan umum, tersedia fasilitas lengkap.
Baik	3	Rumah cukup mudah dijangkau, fasilitas dasar tersedia, tapi belum seideal “sangat baik”.
Kurang Baik	2	Rumah berada di lokasi sulit dijangkau, jauh dari fasilitas umum, akses jalan buruk atau sempit.

2. Kondisi Sosial Ekonomi

Atribut ini menilai tingkat kemampuan ekonomi dan kualitas hidup penghuni, termasuk penghasilan, pekerjaan, pendidikan, dan daya beli.

Partisi	Nilai	Penjelasan
Tinggi	50	Penghuni memiliki penghasilan stabil, pekerjaan tetap, bisa memperbaiki atau membangun rumah dengan baik.

Sedang	30	Penghuni mampu memenuhi kebutuhan pokok dan sebagian kebutuhan sekunder. Kondisi rumah cukup baik, namun terbatas.
Rendah	20	Penghasilan minim, pekerjaan tidak tetap, tidak mampu memperbaiki rumah, rawan tinggal di hunian tak layak.

Tabel 3. 4. Data Sampel Penelitian Sudah Dikonversi

Nama	Jenis Kelamin	Aksesibilitas Infrastruktur	Kondisi Sosial Ekonomi	NIK	Alamat	Tipe Bangunan	Jalan
Adrian Putra	Laki-Laki	5	30	1213076592183740	Padang Matinggi	42	Beton
Bayu Kurniawan	Laki-Laki	3	50	1218927365012640	Padang Matinggi	36	Beton
Candra Wijaya	Laki-Laki	5	50	1211763928471520	Padang Matinggi	42	Beton
Denny Prasetyo	Laki-Laki	3	20	1213647859206170	Padang Matinggi	42	Beton
Evan Saputra	Laki-Laki	2	20	1215782039461580	Padang Matinggi	42	Beton
Farhan Maulana	Laki-Laki	5	30	1219056217483200	Padang Matinggi	42	Beton
Gilang Ramadhani	Laki-Laki	5	50	1214398271059860	Padang Matinggi	42	Beton
Hendra Setiawan	Laki-Laki	5	50	1216283741905270	Padang Matinggi	42	Beton
Iqbal Permana	Laki-Laki	5	30	1213195628471090	Padang Matinggi	42	Beton
Junaidi Akbar	Laki-Laki	5	30	1217946352184060	Padang Matinggi	36	Beton

3.8.3. Perhitungan Data

Untuk tahapan awal melakukan perhitungan pada metode K-Means Clustering yaitu menghitung centroid dari 3 data yang ada pada data sampel. Untuk menghitung data nya juga, yang akan diambil awalnya yaitu data pertama dan

kedua. Untuk proses perhitungan juga terdapat beberapa tahapan yaitu sebagai berikut.

$$d_{1,1} = \sqrt{(titik\ 1\ di\ x1 - data\ 1\ di\ x1)^2 + (titik\ 1\ di\ x2 - data\ 1\ di\ x2)^2 + (titik\ 1\ di\ x3 - data\ 1\ di\ x3)^2}$$

$$d_{1,2} = \sqrt{(titik\ 2\ di\ x1 - data\ 2\ di\ x1)^2 + (titik\ 2\ di\ x2 - data\ 2\ di\ x2)^2 + (titik\ 2\ di\ x3 - data\ 2\ di\ x3)^2}$$

$$d_{1,3} = \sqrt{(titik\ 3\ di\ x1 - data\ 3\ di\ x1)^2 + (titik\ 3\ di\ x2 - data\ 3\ di\ x2)^2 + (titik\ 3\ di\ x3 - data\ 3\ di\ x3)^2}$$

Rumus yang kamu tulis di atas merupakan representasi dari perhitungan jarak Euclidean antara sebuah data dengan titik centroid cluster dalam algoritma K-Means. Jarak ini dihitung dengan mengukur perbedaan antara setiap nilai fitur (dalam hal ini X1, X2, dan X3) dari data terhadap nilai fitur yang sesuai pada centroid cluster. Misalnya, untuk menghitung jarak antara data ke-1 dengan centroid cluster-1 ($d_{1,1}$), kita kurangkan nilai X1 data dengan X1 centroid, lalu kuadratkan hasilnya, dan lakukan hal yang sama untuk X2 dan X3, lalu jumlahkan semuanya dan akarnya diambil (akar kuadrat dari jumlah kuadrat perbedaan). Ini dilakukan untuk semua kombinasi data dengan centroid-cluster lainnya seperti $d_{1,2}$ dan $d_{1,3}$.

Perhitungan jarak ini penting karena menentukan sejauh mana suatu data "dekat" atau "mirip" dengan centroid dari sebuah cluster. Data akan dikelompokkan ke cluster yang memiliki jarak terkecil terhadap centroid-nya. Setelah seluruh data diklasifikasikan ke dalam cluster masing-masing, analisis seperti Average Within

Centroid Distance (AWCD) dapat dilakukan untuk mengukur seberapa padat (kompak) masing-masing cluster. Semakin kecil jaraknya, maka semakin baik kualitas pengelompokannya, karena menunjukkan bahwa anggota dalam cluster tersebut memiliki karakteristik yang lebih mirip satu sama lain.

$$\text{Centroid 1} = 5, 30$$

$$\text{Centroid 2} = 3, 50$$

$$\text{Centroid 3} = 5, 50$$

$$\text{Data 1} = 5, 30$$

$$\begin{aligned} d_{1,1} &= \sqrt{(5 - 5)^2 + (30 - 30)^2} \\ &= \sqrt{(0)^2 + (0)^2} \\ &= \mathbf{0} \end{aligned}$$

$$\begin{aligned} d_{1,2} &= \sqrt{(5 - 3)^2 + (30 - 50)^2} \\ &= \sqrt{(2)^2 + (-20)^2} \\ &= \mathbf{20} \end{aligned}$$

$$\begin{aligned} d_{1,3} &= \sqrt{(5 - 5)^2 + (30 - 50)^2} \\ &= \sqrt{(0)^2 + (-20)^2} \\ &= \mathbf{20} \end{aligned}$$

$$\text{Data 2} = 3, 50$$

$$\begin{aligned} D_{2,1} &= \sqrt{(3 - 5)^2 + (50 - 30)^2} \\ &= \sqrt{(-2)^2 + (20)^2} \\ &= \mathbf{20} \end{aligned}$$

$$\begin{aligned} D_{2,2} &= \sqrt{(3 - 5)^2 + (50 - 50)^2} \\ &= \sqrt{(0)^2 + (0)^2} \\ &= \mathbf{0} \end{aligned}$$

$$\begin{aligned} D_{2,3} &= \sqrt{(3 - 5)^2 + (50 - 50)^2} \\ &= \sqrt{(-2)^2 + (0)^2} \\ &= \mathbf{2} \end{aligned}$$

Data 3 = 5, 50

$$\begin{aligned} d_{3,1} &= \sqrt{(5 - 5)^2 + (30 - 50)^2} \\ &= \sqrt{(0)^2 + (-20)^2} \\ &= \mathbf{20} \end{aligned}$$

$$\begin{aligned} d_{3,2} &= \sqrt{(5 - 3)^2 + (50 - 50)^2} \\ &= \sqrt{(2)^2 + (0)^2} \\ &= \mathbf{2} \end{aligned}$$

$$\begin{aligned} d_{3,3} &= \sqrt{(5 - 5)^2 + (50 - 50)^2} \\ &= \sqrt{(0)^2 + (0)^2} \\ &= \mathbf{0} \end{aligned}$$

Cluster	Centroid			
	X1	X2	X3	Kelompok
C1	0	20	20	1
C2	20	0	2	2
C3	20	2	0	3

Data 4 = 3, 20

$$\begin{aligned} d_{4,1} &= \sqrt{(3 - 5)^2 + (20 - 30)^2} \\ &= \sqrt{(-2)^2 + (-10)^2} \\ &= \mathbf{10} \end{aligned}$$

$$\begin{aligned} d_{4,2} &= \sqrt{(3 - 3)^2 + (20 - 50)^2} \\ &= \sqrt{(0)^2 + (-30)^2} \\ &= \mathbf{30} \end{aligned}$$

$$\begin{aligned} d_{4,3} &= \sqrt{(3 - 5)^2 + (20 - 50)^2} \\ &= \sqrt{(-2)^2 + (-30)^2} \\ &= \mathbf{30} \end{aligned}$$

Cluster	Centroid			
	X1	X2	X3	Kelompok
C4	10	30	30	1

Update Centroid

Centroid	X1	X2
d1	$(5+3)/2 = 4$	$(30+20)/2 = 25$

Data 5 = 2, 20

$$\begin{aligned}
 d_{5,1} &= \sqrt{(2-4)^2 + (20-25)^2} \\
 &= \sqrt{(-2)^2 + (-5)^2} \\
 &= \mathbf{5}
 \end{aligned}$$

$$\begin{aligned}
 d_{5,2} &= \sqrt{(2-3)^2 + (20-50)^2} \\
 &= \sqrt{(-1)^2 + (-30)^2} \\
 &= \mathbf{30}
 \end{aligned}$$

$$\begin{aligned}
 d_{5,3} &= \sqrt{(2-5)^2 + (20-50)^2} \\
 &= \sqrt{(-3)^2 + (-30)^2} \\
 &= \mathbf{30}
 \end{aligned}$$

Cluster	Centroid			
	X1	X2	X3	Kelompok
C5	5	30	30	1

Update Centroid

Centroid	X1	X2
d1	$(4+2)/2 = 3$	$(20+25)/2 = 22$

Data 6 = 5, 30

$$\begin{aligned}
 d_{6,1} &= \sqrt{(5-3)^2 + (30-22)^2} \\
 &= \sqrt{(2)^2 + (8)^2} \\
 &= \mathbf{8}
 \end{aligned}$$

$$\begin{aligned}
 d_{6,2} &= \sqrt{(5-3)^2 + (30-50)^2} \\
 &= \sqrt{(2)^2 + (-20)^2} \\
 &= \mathbf{20}
 \end{aligned}$$

$$\begin{aligned}
 d_{6,3} &= \sqrt{(5 - 5)^2 + (30 - 50)^2} \\
 &= \sqrt{(0)^2 + (-20)^2} \\
 &= \mathbf{20}
 \end{aligned}$$

Cluster	Centroid			
	X1	X2	X3	Kelompok
C6	8	20	20	1

Update Centroid

Centroid	X1	X2
d1	$(3+5)/2 = 4$	$(22+30)/2 = 26$

Data 7 = 5, 50

$$\begin{aligned}
 d_{7,1} &= \sqrt{(5 - 4)^2 + (50 - 26)^2} \\
 &= \sqrt{(1)^2 + (24)^2} \\
 &= \mathbf{24}
 \end{aligned}$$

$$\begin{aligned}
 d_{7,2} &= \sqrt{(5 - 3)^2 + (50 - 50)^2} \\
 &= \sqrt{(2)^2 + (0)^2} \\
 &= \mathbf{2}
 \end{aligned}$$

$$\begin{aligned}
 d_{7,3} &= \sqrt{(5 - 5)^2 + (50 - 50)^2} \\
 &= \sqrt{(0)^2 + (0)^2} \\
 &= \mathbf{0}
 \end{aligned}$$

Cluster	Centroid			
	X1	X2	X3	Kelompok
C7	24	2	0	3

Update Centroid

Centroid	X1	X2
d3	$(5+5)/2 = 5$	$(50+50)/2 = 50$

Data 8 = 5, 50

$$d_{8,1} = \sqrt{(5 - 4)^2 + (50 - 26)^2}$$

$$= \sqrt{(1)^2 + (24)^2}$$

$$= \mathbf{24}$$

$$d_{8,2} = \sqrt{(5-3)^2 + (50-50)^2}$$

$$= \sqrt{(2)^2 + (0)^2}$$

$$= \mathbf{2}$$

$$d_{8,3} = \sqrt{(5-5)^2 + (50-50)^2}$$

$$= \sqrt{(0)^2 + (0)^2}$$

$$= \mathbf{0}$$

Cluster	Centroid			
	X1	X2	X3	Kelompok
C8	24	2	0	3

Update Centroid

Centroid	X1	X2
d3	$(5+5)/2 = 5$	$(50+50)/2 = 50$

Data 9 = 5, 30

$$d_{9,1} = \sqrt{(5-4)^2 + (30-26)^2}$$

$$= \sqrt{(1)^2 + (4)^2}$$

$$= \mathbf{4}$$

$$d_{9,2} = \sqrt{(5-3)^2 + (30-50)^2}$$

$$= \sqrt{(2)^2 + (-20)^2}$$

$$= \mathbf{20}$$

$$d_{9,3} = \sqrt{(5-5)^2 + (30-50)^2}$$

$$= \sqrt{(0)^2 + (-20)^2}$$

$$= \mathbf{20}$$

Cluster	Centroid			
	X1	X2	X3	Kelompok
C9	4	20	20	1

Update Centroid

Centroid	X1	X2
d1	$(4+5)/2 = 4$	$(26+30)/2 = 28$

Data 10 = 5, 30

$$\begin{aligned} d_{10,1} &= \sqrt{(5 - 4)^2 + (30 - 28)^2} \\ &= \sqrt{(1)^2 + (2)^2} \\ &= 2 \end{aligned}$$

$$\begin{aligned} d_{10,2} &= \sqrt{(5 - 3)^2 + (30 - 50)^2} \\ &= \sqrt{(2)^2 + (-20)^2} \\ &= 20 \end{aligned}$$

$$\begin{aligned} d_{10,3} &= \sqrt{(5 - 5)^2 + (30 - 50)^2} \\ &= \sqrt{(0)^2 + (-20)^2} \\ &= 20 \end{aligned}$$

Cluster	Centroid			
	X1	X2	X3	Kelompok
C10	2	20	20	1

Hasil Clustering

Tabel 3. 2. Hasil Cluster menggunakan Metode K-Means

No	Nama	(X1)	(X2)	(X3)	Kelompok
1	Adrian Putra	0	20	20	1
2	Bayu Kurniawan	20	0	2	2
3	Candra Wijaya	20	2	0	3
4	Denny Prasetyo	10	30	30	1
5	Evan Saputra	5	30	30	1
6	Farhan Maulana	8	20	20	1
7	Gilang Ramadhan	24	2	0	3
8	Hendra Setiawan	24	2	0	3
9	Iqbal Permana	4	20	20	1
10	Junaidi Akbar	2	20	20	1

Pada penelitian ini, metode K-Means Clustering telah diterapkan untuk mengelompokkan kawasan permukiman berdasarkan tiga variabel utama, yaitu infrastruktur dasar (X1), aksesibilitas (X2), dan kondisi sosial ekonomi (X3). Proses klasterisasi menghasilkan tiga kelompok atau Cluster yang masing-masing

menunjukkan karakteristik hunian yang relatif homogen. Metode ini efektif dalam menyederhanakan kompleksitas data dan memberikan representasi visual dan numerik mengenai kondisi kawasan permukiman yang tersebar.

Dari hasil klasterisasi, diperoleh bahwa Cluster 1 (C1) terdiri dari kawasan dengan aksesibilitas sangat baik dan kondisi sosial ekonomi tinggi, meskipun nilai infrastruktur tidak sepenuhnya maksimal. Kawasan dalam C1 dikategorikan sebagai layak huni, karena dua dari tiga indikator utama berada pada kategori unggul, sehingga mampu mendukung kualitas hidup yang baik. Sementara itu, Cluster 2 (C2) menunjukkan dominasi nilai tinggi pada aspek infrastruktur dasar, namun memiliki nilai rendah pada aksesibilitas dan kondisi sosial ekonomi. Karena ketidakseimbangan aspek ini, kawasan dalam C2 dapat dikategorikan sebagai kurang layak huni. Adapun Cluster 3 (C3) menggambarkan kawasan yang hanya unggul pada infrastruktur, namun sangat rendah dalam aspek sosial ekonomi, sehingga dikategorikan sebagai tidak layak huni secara menyeluruh.

Dengan adanya segmentasi seperti ini, Dinas Perumahan dan Kawasan Permukiman dapat memperoleh gambaran yang lebih tajam tentang kualitas lingkungan permukiman berdasarkan indikator yang telah dipetakan. Kawasan dalam C1 dapat dijadikan model atau prioritas pelestarian, sementara kawasan dalam C2 dan C3 dapat menjadi fokus program intervensi, baik dalam bentuk pembangunan sarana prasarana maupun peningkatan kesejahteraan sosial ekonomi warga. Dengan pendekatan berbasis data ini, kebijakan pembangunan dapat disusun secara lebih terarah, efisien, dan responsif terhadap kebutuhan nyata di lapangan.