

BAB IV

HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Pengumpulan data dalam penelitian ini merupakan hasil dari proses pengumpulan data primer dan atau sekunder yang berisi informasi terkait karakteristik kawasan permukiman berdasarkan dua indikator utama, yaitu aksesibilitas infrastruktur dan kondisi sosial ekonomi. Data ini dihimpun dengan tujuan untuk mendukung analisis kelayakan hunian pada masing-masing kawasan yang dihuni oleh individu-individu dengan identitas tertentu. Kategori seperti “Sangat Baik,” “Baik,” dan “Kurang Baik” untuk aksesibilitas serta “Tinggi,” “Sedang,” dan “Rendah” untuk kondisi sosial ekonomi diperoleh melalui pengamatan langsung, kuesioner, atau dokumentasi dari Dinas Perumahan dan Kawasan Permukiman. Informasi ini penting sebagai dasar untuk proses *Clusterisasi* menggunakan algoritma K-Means, agar dapat mengelompokkan kawasan dengan karakteristik yang serupa secara sistematis dan objektif guna mendukung pengambilan keputusan dalam perencanaan dan pengembangan kawasan permukiman yang lebih tepat sasaran.

Tabel 4. 1. Data Sampel

Nama	NIK	Alamat	Tipe Bangunan	Jalan	Aksesibilitas Infrastruktur	Kondisi Sosial Ekonomi
Adrian Akbar	1219018273621	Padang Matinggi	Permanen	Beton	Kurang Baik	Tinggi
Bagus Kurniawan	1214562317890	Padang Matinggi	Semi Permanen	Beton	Baik	Rendah
Bagus Nugraha	1219837456120	Padang Matinggi	Permanen	Beton	Baik	Tinggi
Bagus Ramadhan	1217423890654	Padang Matinggi	Permanen	Beton	Baik	Sedang
Bagus Wibowo	1216138409257	Padang Matinggi	Semi Permanen	Beton	Baik	Sedang
Bayu Nugraha	1215028493178	Padang Matinggi	Permanen	Beton	Baik	Sedang
Candra Akbar	1217683021549	Padang Matinggi	Semi Permanen	Beton	Kurang Baik	Tinggi
Candra Permana	1213847209612	Padang Matinggi	Permanen	Beton	Baik	Tinggi
Denny Prasetyo	1216918574302	Padang Matinggi	Permanen	Beton	Sangat Baik	Rendah
Dimas Akbar	1215749208364	Padang Matinggi	Permanen	Beton	Sangat Baik	Sedang
Dimas Anwar	1218057391620	Padang Matinggi	Semi Permanen	Beton	Baik	Sedang
Dimas Hidayat	1213420857169	Padang Matinggi	Permanen	Beton	Sangat Baik	Rendah
Dimas Putra	1216798124037	Padang Matinggi	Semi Permanen	Beton	Kurang Baik	Sedang
Dimas Riansyah	1219304812650	Padang Matinggi	Sementara	Beton	Sangat Baik	Rendah
Eko Kurniawan	1217439086214	Padang Matinggi	Sementara	Beton	Kurang Baik	Rendah
Eko Prasetyo	1211957208463	Padang Matinggi	Permanen	Beton	Sangat Baik	Tinggi
Farhan Fauzan	1216382940759	Padang Matinggi	Semi Permanen	Beton	Kurang Baik	Sedang

Farhan Kurniawan	1217903846215	Padang Matinggi	Permanen	Beton	Baik	Sedang
Farhan Nugraha	1212348907612	Padang Matinggi	Sementara	Beton	Baik	Rendah
Farhan Permana	1216849301723	Padang Matinggi	Semi Permanen	Beton	Kurang Baik	Rendah
Farhan Wibowo	1215062748913	Padang Matinggi	Permanen	Beton	Kurang Baik	Tinggi
Gerry Akbar	1216789052319	Padang Matinggi	Permanen	Beton	Baik	Sedang
Gerry Gunawan	1218046752193	Padang Matinggi	Semi Permanen	Beton	Sangat Baik	Rendah
Gilang Akbar	1219328071543	Padang Matinggi	Permanen	Beton	Sangat Baik	Tinggi
Gilang Akbar	1214597038214	Padang Matinggi	Permanen	Beton	Baik	Tinggi
Gilang Fauzan	1217601943857	Padang Matinggi	Semi Permanen	Beton	Baik	Rendah
Gilang Saputra	1216978204306	Padang Matinggi	Semi Permanen	Beton	Sangat Baik	Rendah
Gilang Wijaya	1218239071568	Padang Matinggi	Sementara	Beton	Baik	Rendah
Hendra Anwar	1215187402936	Padang Matinggi	Semi Permanen	Beton	Kurang Baik	Sedang
Hendra Setiawan	1218097245163	Padang Matinggi	Permanen	Beton	Sangat Baik	Tinggi
Iqbal Permana	1219307152834	Padang Matinggi	Permanen	Beton	Baik	Sedang
Junaidi Akbar	1216073149286	Padang Matinggi	Semi Permanen	Beton	Baik	Sedang
Kadek Wijaya	1218406729153	Padang Matinggi	Sementara	Beton	Kurang Baik	Rendah
Kiki Ramadhan	1219847032147	Padang Matinggi	Permanen	Beton	Sangat Baik	Tinggi
Lazuardi Putra	1215793208645	Padang Matinggi	Semi Permanen	Beton	Kurang Baik	Rendah
M. Fikri Akbar	1212937405861	Padang Matinggi	Permanen	Beton	Baik	Sedang

M. Ghozali	1217683209543	Padang Matinggi	Permanen	Beton	Sangat Baik	Sedang
M. Iqbal Hidayat	1214598732064	Padang Matinggi	Semi Permanen	Beton	Baik	Rendah
M. Ramadhan	1216930487152	Padang Matinggi	Sementara	Beton	Kurang Baik	Rendah
M. Rizki Kurniawan	1219078352610	Padang Matinggi	Permanen	Beton	Baik	Sedang
Naufal Prasetya	1213847691052	Padang Matinggi	Semi Permanen	Beton	Sangat Baik	Rendah
Naufal Ramadhan	1215098372061	Padang Matinggi	Permanen	Beton	Baik	Sedang
Naufal Yusron	1217604981320	Padang Matinggi	Permanen	Beton	Sangat Baik	Tinggi
Nazaruddin Anwar	1216743820591	Padang Matinggi	Sementara	Beton	Kurang Baik	Rendah
Oka Permana	1218907246135	Padang Matinggi	Semi Permanen	Beton	Baik	Rendah
Pandu Wicaksono	1216058372196	Padang Matinggi	Permanen	Beton	Baik	Sedang
Rangga Putra	1219803472056	Padang Matinggi	Semi Permanen	Beton	Kurang Baik	Rendah
Reza Hidayat	1217590138402	Padang Matinggi	Permanen	Beton	Sangat Baik	Tinggi
Reza Kurniawan	1213420891765	Padang Matinggi	Semi Permanen	Beton	Kurang Baik	Rendah
Reza Permana	1217869304720	Padang Matinggi	Permanen	Beton	Baik	Tinggi
Rio Saputra	1219038742165	Padang Matinggi	Sementara	Beton	Kurang Baik	Sedang
Rizki Gunawan	1216985740312	Padang Matinggi	Semi Permanen	Beton	Baik	Rendah
Rizki Saputra	1215748309640	Padang Matinggi	Permanen	Beton	Sangat Baik	Sedang
Rizky Hidayat	1214389201765	Padang Matinggi	Sementara	Beton	Kurang Baik	Rendah
Roni Pratama	1218604279153	Padang Matinggi	Permanen	Beton	Baik	Tinggi

Sandi Prasetyo	1219028473159	Padang Matinggi	Semi Permanen	Beton	Baik	Sedang
Satria Nugraha	1216840291574	Padang Matinggi	Permanen	Beton	Sangat Baik	Tinggi
Surya Akbar	1213098472057	Padang Matinggi	Semi Permanen	Beton	Kurang Baik	Rendah
Wahyu Gunawan	1215748930271	Padang Matinggi	Permanen	Beton	Baik	Sedang
Yudi Pranata	1217409826310	Padang Matinggi	Semi Permanen	Beton	Baik	Rendah

Pada tabel di atas merupakan contoh data sampel penelitian yang akan digunakan dalam proses analisis pada penelitian ini. Tabel tersebut memuat informasi mengenai nama responden, jenis kelamin, aksesibilitas infrastruktur, dan kondisi sosial ekonomi yang menjadi variabel utama dalam penelitian. Data yang ditampilkan merupakan bagian dari keseluruhan data yang akan dianalisis, di mana total jumlah data sampel yang digunakan dalam penelitian ini adalah sebanyak 60 data. Setiap sampel mewakili kondisi aktual dari lingkungan permukiman yang akan diklasifikasikan berdasarkan tingkat kelayakan huniannya. Data ini nantinya akan diolah menggunakan algoritma K-Means untuk membentuk *Cluster-Cluster* yang menggambarkan karakteristik masing-masing kawasan permukiman.

4.2. Preprocessing Data

Tahapan preprocessing data merupakan proses penting dalam penelitian ini yang bertujuan untuk mempersiapkan data agar siap dianalisis, termasuk menyaring, menyeleksi, dan memastikan kesesuaian format. Meskipun data yang diperoleh sudah tersusun dengan baik, lengkap, dan rapi, sehingga tidak memerlukan banyak proses pembersihan atau perbaikan, tetap dilakukan langkah penting berupa transformasi data. Pada tahap ini, data kategorikal diubah menjadi

bentuk numerikal agar dapat diproses oleh algoritma K-Means, mengingat beberapa algoritma hanya dapat menangani atribut numerik. Dengan demikian, preprocessing lebih berfokus pada konversi format data agar sesuai dengan kebutuhan analisis. Tetapi untuk merancang model cluster ini dengan menggunakan data diatas, data akan di ubah dulu dalam bentuk nimeril, 0, 1, dan 2. Hal ini dilakukan agar data dapat diolah pada aplikasi rapidminer. Untuk perubahan datanya yaitu sebagai berikut.

Atribut	Partisi	Nilai
Akseibilitas Infrastruktur	Sangat Baik	2
	Baik	1
	Kurang Baik	0
Kondisi Sosial Ekonomi	Tinggi	2
	Sedang	1
	Rendah	0

Dalam proses pengolahan data menggunakan metode K-Means clustering pada aplikasi RapidMiner, seluruh data kategorikal perlu diubah ke dalam bentuk numerik agar dapat diproses secara optimal oleh algoritma. Oleh karena itu, atribut seperti Aksesibilitas Infrastruktur dan Kondisi Sosial Ekonomi yang semula berbentuk teks (Sangat Baik, Baik, Kurang Baik serta Tinggi, Sedang, Rendah) dikonversi ke bentuk angka dengan skala ordinal, yaitu 2 untuk kategori tertinggi, 1 untuk kategori sedang, dan 0 untuk kategori terendah. Konversi ini dilakukan karena algoritma K-Means hanya dapat menghitung jarak antar data jika semua atribut bersifat numerik.

Penjelasan Atribut dan Partisi

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	2	Rumah berada di lokasi strategis, dekat jalan utama, mudah dijangkau kendaraan umum, tersedia fasilitas lengkap.
Baik	1	Rumah cukup mudah dijangkau, fasilitas dasar tersedia, tapi belum seideal “sangat baik”.
Kurang Baik	0	Rumah berada di lokasi sulit dijangkau, jauh dari fasilitas umum, akses jalan buruk atau sempit.

Kondisi Sosial Ekonomi

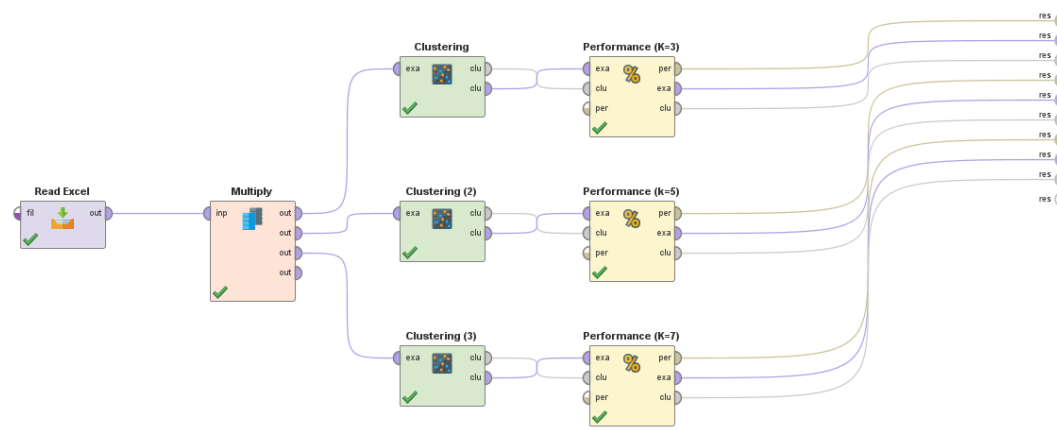
Atribut ini menilai tingkat kemampuan ekonomi dan kualitas hidup penghuni, termasuk penghasilan, pekerjaan, pendidikan, dan daya beli.

Partisi	Nilai	Penjelasan
Tinggi	2	Penghuni memiliki penghasilan stabil, pekerjaan tetap, bisa memperbaiki atau membangun rumah dengan baik.
Sedang	1	Penghuni mampu memenuhi kebutuhan pokok dan sebagian kebutuhan sekunder. Kondisi rumah cukup baik, namun terbatas.
Rendah	0	Penghasilan minim, pekerjaan tidak tetap, tidak mampu memperbaiki rumah, rawan tinggal di hunian tak layak.

4.3. Perancangan Model Cluster

Perancangan model *Cluster* merupakan tahap inti dalam penelitian ini yang bertujuan untuk membentuk kelompok-kelompok data berdasarkan kemiripan karakteristik menggunakan algoritma K-Means. Pada tahap ini, variabel-variabel seperti Aksesibilitas Infrastruktur dan Kondisi Sosial Ekonomi dijadikan sebagai atribut utama untuk membentuk *Cluster* yang merepresentasikan tingkat kelayakan hunian. Proses ini diawali dengan penentuan jumlah *Cluster* optimal, kemudian dilanjutkan dengan perhitungan jarak antar data dan pusat *Cluster* (centroid), hingga tercapai konvergensi atau tidak ada lagi perubahan signifikan dalam pengelompokan. Dengan merancang model *Cluster* yang tepat, hasil

pengelompokan dapat mencerminkan kondisi riil di lapangan dan menjadi dasar yang kuat untuk pengambilan keputusan oleh instansi terkait.



Gambar 4. 1. Perancangan Model Cluster

Gambar di atas merupakan diagram alur perancangan model klasterisasi menggunakan metode K-Means pada aplikasi RapidMiner. Proses dimulai dengan membaca data dari file Excel melalui operator Read Excel, kemudian data digandakan tiga kali menggunakan Multiply agar dapat digunakan secara paralel dalam tiga konfigurasi klaster yang berbeda. Masing-masing cabang dari proses ini diteruskan ke operator Clustering dengan parameter jumlah klaster yang berbeda: K=3, K=5, dan K=7, untuk mengeksplorasi jumlah klaster optimal. Hasil dari tiap proses klasterisasi kemudian dianalisis menggunakan operator Performance guna mengevaluasi kualitas klaster berdasarkan metrik seperti *cluster compactness* dan *separation*. Pendekatan ini memungkinkan perbandingan hasil klasterisasi untuk menentukan nilai K yang paling sesuai terhadap pola dalam data.

4.4. Hasil Cluster

Hasil cluster yang diperoleh dari proses analisis menggunakan algoritma K-Means akan disajikan dalam bentuk tabel guna mempermudah interpretasi terhadap pola pengelompokan data. Setiap baris dalam tabel akan menampilkan informasi

data beserta label cluster sebagai hasil akhir dari proses pengelompokan. Dalam penelitian ini, hasil yang ditampilkan mencakup tiga skenario jumlah cluster, yaitu $K = 3$, $K = 5$, dan $K = 7$. Penyajian dalam beberapa skenario ini bertujuan untuk memberikan perbandingan dan evaluasi terhadap variasi jumlah cluster, sehingga dapat diketahui jumlah cluster mana yang paling optimal dalam merepresentasikan karakteristik data yang dianalisis.

Tabel 4. 2. Hasil Cluster

Nama	Jenis Kelamin	Aksesibilitas Infrastruktur	Kondisi Sosial Ekonomi	K (3)	K (5)	K (7)
Adrian Akbar	Laki-Laki	Kurang Baik	Tinggi	1	1	3
Bagus Kurniawan	Laki-Laki	Cukup	Rendah	2	5	7
Bagus Nugraha	Laki-Laki	Cukup	Tinggi	1	1	5
Bagus Ramadhan	Laki-Laki	Cukup	Sedang	1	3	4
Bagus Wibowo	Laki-Laki	Cukup	Sedang	1	3	4
Bayu Nugraha	Laki-Laki	Cukup	Sedang	1	3	4
Candra Akbar	Laki-Laki	Kurang Baik	Tinggi	1	1	3
Candra Permana	Laki-Laki	Cukup	Tinggi	1	1	5
Denny Prasetyo	Laki-Laki	Baik	Rendah	2	5	2
Dimas Akbar	Laki-Laki	Baik	Sedang	1	2	2
Dimas Anwar	Laki-Laki	Cukup	Sedang	1	3	4
Dimas Hidayat	Laki-Laki	Baik	Rendah	2	5	2
Dimas Putra	Laki-Laki	Kurang Baik	Sedang	3	4	3
Dimas Riansyah	Laki-Laki	Baik	Rendah	2	5	2
Eko Kurniawan	Laki-Laki	Kurang Baik	Rendah	3	4	6
Eko Prasetyo	Laki-Laki	Baik	Tinggi	1	2	1
Farhan Fauzan	Laki-Laki	Kurang Baik	Sedang	3	4	3
Farhan Kurniawan	Laki-Laki	Cukup	Sedang	1	3	4
Farhan Nugraha	Laki-Laki	Cukup	Rendah	2	5	7
Farhan Permana	Laki-Laki	Kurang Baik	Rendah	3	4	6
Farhan Wibowo	Laki-Laki	Kurang Baik	Tinggi	1	1	3
Gerry Akbar	Laki-Laki	Cukup	Sedang	1	3	4
Gerry Gunawan	Laki-Laki	Baik	Rendah	2	5	2
Gilang Akbar	Laki-Laki	Baik	Tinggi	1	2	1
Gilang Akbar	Laki-Laki	Cukup	Tinggi	1	1	5
Gilang Fauzan	Laki-Laki	Cukup	Rendah	2	5	7
Gilang Saputra	Laki-Laki	Baik	Rendah	2	5	2
Gilang Wijaya	Laki-Laki	Cukup	Rendah	2	5	7
Hendra Anwar	Laki-Laki	Kurang Baik	Sedang	3	4	3
Hendra Setiawan	Laki-Laki	Baik	Tinggi	1	2	1
Iqbal Setiawan	Laki-Laki	Baik	Sedang	1	3	4
Junaidi Akbar	Laki-Laki	Kurang Baik	Sedang	1	3	4
Kevin Akbar	Laki-Laki	Cukup	Tinggi	3	4	6
Kevin Halim	Laki-Laki	Baik	Rendah	1	2	1
Lutfi Kurniawan	Laki-Laki	Kurang Baik	Sedang	3	4	6
Lutfi Permana	Laki-Laki	Baik	Tinggi	1	3	4

Lutfi Setiawan	Laki-Laki	Kurang Baik	Rendah	1	2	2
Maulana Maulana	Laki-Laki	Cukup	Tinggi	2	5	7
Niko Permana	Laki-Laki	Cukup	Rendah	3	4	6
Niko Putra	Laki-Laki	Kurang Baik	Sedang	1	3	4
Omar Santoso	Laki-Laki	Cukup	Sedang	2	5	2
Prasetyo Fauzan	Laki-Laki	Kurang Baik	Tinggi	1	3	4
Prasetyo Gunawan	Laki-Laki	Kurang Baik	Rendah	1	2	1
Prasetyo Prasetyo	Laki-Laki	Baik	Rendah	3	4	6
Qomar Halim	Laki-Laki	Cukup	Sedang	2	5	7
Rizky Santoso	Laki-Laki	Baik	Rendah	1	3	4
Syahrul Hidayat	Laki-Laki	Kurang Baik	Tinggi	3	4	6
Taufik Hidayat	Laki-Laki	Baik	Rendah	1	2	1
Taufik Maulana	Laki-Laki	Baik	Rendah	3	4	6
Umar Riansyah	Laki-Laki	Cukup	Rendah	1	1	5
Umar Riansyah	Laki-Laki	Baik	Tinggi	3	4	3
Umar Wibowo	Laki-Laki	Baik	Tinggi	2	5	7
Wahyu Santoso	Laki-Laki	Baik	Tinggi	1	2	2
Yuda Anwar	Laki-Laki	Baik	Tinggi	3	4	6
Yuda Halim	Laki-Laki	Cukup	Tinggi	1	1	5
Yuda Nasution	Laki-Laki	Kurang Baik	Rendah	1	3	4
Yuda Nugraha	Laki-Laki	Kurang Baik	Rendah	1	2	1
Yuda Wijaya	Laki-Laki	Baik	Sedang	3	4	6
Zaky Maulana	Laki-Laki	Baik	Rendah	1	3	4
Zaky Ramadhan	Laki-Laki	Kurang Baik	Sedang	2	5	7

Pada tabel di atas merupakan hasil dari proses clustering terhadap 60 data responden berdasarkan tiga variabel utama, yaitu Aksesibilitas Infrastruktur, Kondisi Sosial Ekonomi, dan Jenis Kelamin (walau seluruh responden adalah laki-laki). Pengelompokan dilakukan dengan algoritma K-Means, dan hasilnya dianalisis dalam tiga skenario jumlah cluster yang berbeda: $K = 3$, $K = 5$, dan $K = 7$. Setiap skenario memberikan gambaran yang berbeda mengenai bagaimana data tersebut dikelompokkan berdasarkan kesamaan karakteristik.

Ketika data dikelompokkan menjadi 3 klaster, terlihat adanya 3 tingkatan kondisi bangunan. Klaster 1 yang terdiri dari 39 rumah didominasi oleh kondisi *aksesibilitas cukup* dan *sosial ekonomi sedang hingga tinggi*. Ini bisa dikategorikan sebagai bangunan yang layak huni, walau mungkin tidak sempurna, tetapi sudah memenuhi standar dasar. Klaster 2 dengan 13 rumah mayoritas berada di lingkungan dengan *aksesibilitas baik* tapi *kondisi ekonomi rendah*. Rumah-rumah

ini tampak layak dari sisi infrastruktur, tetapi *mungkin sulit dipertahankan secara ekonomi*. Ini bisa disebut layak tapi rentan. Klaster 3 (8 rumah) cenderung mencerminkan rumah-rumah yang tidak layak huni: *aksesibilitas infrastruktur kurang baik dan kondisi sosial ekonomi rendah atau sedang*. Rumah dalam klaster ini kemungkinan dibangun seadanya dan berada di bagian wilayah yang kurang diperhatikan pembangunan fasilitasnya.

Pada pemetaan menjadi 5 klaster, terlihat pengelompokan yang lebih rinci. Klaster 1 (21 rumah) sangat kuat secara infrastruktur dan ekonomi – bisa dikatakan ini adalah bangunan paling layak huni, kemungkinan dibangun dengan material dan desain yang baik. Klaster 2 (15 rumah) juga termasuk layak, tetapi dengan sedikit keterbatasan di akses. Klaster 3 (13 rumah) terlihat seperti *wilayah transisi*, yaitu rumah-rumah yang secara ekonomi menengah namun infrastrukturnya kurang, sehingga disebut setengah layak huni. Klaster 4 (7 rumah) tergolong tidak layak huni, karena umumnya berada di lokasi dengan infrastruktur yang minim dan ekonomi rendah. Klaster 5 (4 rumah) adalah rumah-rumah yang mungkin secara ekonomi kuat, tetapi berada di daerah dengan tantangan infrastruktur – ini bisa disebut layak dari sisi ekonomi, tidak dari sisi teknis.

Dengan 7 klaster, hasil analisis lebih mendalam lagi. Klaster 1 (9 rumah) adalah rumah-rumah elit dengan infrastruktur baik dan kondisi ekonomi tinggi – sangat layak huni. Klaster 2 (11 rumah) adalah rumah layak huni standar, baik dari sisi lokasi maupun kemampuan penghuni. Klaster 3 (10 rumah) terlihat sebagai rumah dengan ekonomi tinggi tapi akses buruk – bisa disebut *mewah tapi tersembunyi*. Klaster 4 (8 rumah) adalah bangunan tidak layak huni – akses buruk dan ekonomi rendah. Klaster 5 (7 rumah) tergolong miskin secara ekonomi tetapi

berada di lingkungan bagus, mungkin rumah sewa atau rumah bantuan. Klaster 6 (9 rumah) adalah kelompok rumah rata-rata atau sederhana – cukup akses dan ekonomi, layak minimum. Klaster 7 (6 rumah) adalah rumah dengan *ciri sosial ekonomi sedang tapi infrastruktur sangat terbatas*, mereka bisa terancam tidak layak jika tidak dibenahi.

4.5. Hasil Evaluasi

Hasil evaluasi dari proses *clustering* menggunakan algoritma K-Means ditunjukkan melalui nilai Average Within Centroid Distance, yaitu rata-rata jarak antara setiap data dengan pusat cluster-nya masing-masing. Semakin kecil nilai ini, maka semakin baik kualitas pengelompokan karena menunjukkan bahwa data dalam suatu cluster memiliki kesamaan yang tinggi atau lebih terpusat di sekitar centroid-nya. Evaluasi ini membantu dalam menilai seberapa optimal jumlah cluster yang digunakan, sehingga dapat dibandingkan antara $K = 3$, $K = 5$, dan $K = 7$ untuk menentukan mana yang menghasilkan pengelompokan paling konsisten dan representatif terhadap struktur data sebenarnya.

4.5.1. Hasil Avg. Within Centroid Distance

Average Within Centroid Distance adalah ukuran yang menunjukkan rata-rata jarak antara setiap titik data dalam suatu *Cluster* dengan pusat *Cluster* (centroid)-

nya. Nilai ini digunakan untuk menilai seberapa rapat data dalam *Cluster*; semakin kecil nilainya, semakin baik kekompakan *Cluster* tersebut.

Avg. within centroid distance Avg. within centroid distance: -0.429
Avg. within centroid distance Avg. within centroid distance: -0.171
Avg. within centroid distance Avg. within centroid distance: -0.062

Gambar 4. 2. Hasil Evaluasi

Pada gambar di atas ditampilkan hasil evaluasi performa clustering menggunakan metrik Avg. within centroid distance (rata-rata jarak ke pusat klaster), masing-masing dengan nilai negatif yaitu -0.429, -0.171, dan -0.062. Nilai-nilai ini umumnya diperoleh dari proses evaluasi clustering, biasanya ketika kita membandingkan beberapa skenario pengelompokan, seperti jumlah klaster berbeda, algoritma berbeda, atau parameter berbeda dalam metode clustering (misalnya K-Means, Hierarchical, DBSCAN, dsb). Meski umumnya metrik ini bernilai positif (karena jarak tidak bisa negatif secara matematis), nilai negatif di sini bisa jadi karena proses normalisasi/skaling tertentu (misalnya menggunakan cosine similarity atau metrik lain yang dinormalisasi dalam range negatif ke positif).

Secara konsep, Average Within Centroid Distance (AWCD) adalah ukuran seberapa dekat anggota-anggota dalam satu klaster terhadap pusat klasternya

(centroid). Metrik ini sangat penting karena menunjukkan seberapa kompak klaster yang terbentuk. Semakin rendah nilai AWCD (mendekati 0), maka semakin baik hasil klasterisasi tersebut karena anggota klaster berada dekat dengan centroid-nya. Ketika nilai tersebut menjadi negatif (seperti pada gambar), besar kemungkinan itu hasil dari pengukuran dalam skala similarity (kemiripan), bukan murni Euclidean distance. Misalnya, dalam cosine similarity, semakin mendekati 0 atau nilai minimum (tergantung arah skala), menunjukkan kemiripan atau kedekatan yang lebih tinggi.

Jika kita bandingkan tiga hasil tersebut:

1. AWCD = -0.429 menunjukkan rata-rata jarak antar data ke centroid dalam klaster tersebut cukup "jauh" atau "tidak kompak".
2. AWCD = -0.171 berarti klaster mulai lebih rapat atau anggotanya lebih dekat ke centroid.
3. AWCD = -0.062 menunjukkan hasil yang paling baik di antara ketiganya, karena nilainya paling mendekati 0 (atau paling kecil secara absolut dalam skala negatif), menandakan klaster yang paling kompak atau paling optimal.

Dari hasil ini, kita bisa menyimpulkan bahwa model klaster dengan nilai AWCD = -0.062 merupakan yang paling bagus secara struktur karena distribusi anggotanya paling dekat dengan centroid, artinya lebih seragam dan jelas pengelompokannya. Sebaliknya, hasil dengan nilai AWCD lebih besar secara absolut (misalnya -0.429) mungkin menunjukkan adanya data yang tersebar terlalu jauh dari pusat klaster, sehingga mengindikasikan klaster yang kurang optimal. Dalam analisis data, perbandingan seperti ini sangat penting untuk memilih model

klaster terbaik karena dapat memberikan gambaran seberapa efektif suatu metode dalam mengelompokkan data ke dalam kelompok yang homogen. Salah satu indikator yang umum digunakan adalah nilai Average Within Centroid Distance (AWCD), yang mengukur rata-rata jarak antara setiap data dalam suatu cluster dengan centroid-nya. Semakin kecil nilai AWCD, maka semakin padat dan terpusat data dalam cluster tersebut, menandakan bahwa model klasterisasi tersebut memiliki performa yang baik. Dengan melakukan perbandingan AWCD antar cluster atau antar iterasi model, peneliti dapat menentukan struktur klaster yang paling representatif terhadap data, sekaligus memastikan bahwa hasil pengelompokan yang diperoleh benar-benar mencerminkan karakteristik asli dari masing-masing kelompok dalam data tersebut. Oleh karena itu, analisis terhadap hasil klasterisasi tidak hanya berhenti pada jumlah cluster yang terbentuk, tetapi juga mencakup evaluasi terhadap kualitas masing-masing cluster menggunakan ukuran-ukuran seperti AWCD.