

PENERAPAN DATA MINING PADA *CLUSTERING* DATA HARGA RUMAH DKI JAKARTA MENGGUNAKAN ALGORITMA K-MEANS

Teneshya Lidia Putri¹, Raditya Danar Dana²

¹⁾ Komputerisasi Akuntansi STMIK IKMI Cirebon

²⁾ Manajemen Informatika STMIK IKMI Cirebon

Jl Perjuangan No 10B Majasem Kesambi Kota Cirebon

neshyaputri41@gmail.com

ABSTRAK

Rumah merupakan kebutuhan primer manusia, dan data harga rumah menjadi krusial untuk memahami pasar perumahan. Data harga rumah di setiap daerah berbeda-beda sesuai dengan daerah dan kategorinya masing-masing, khususnya pada data harga rumah yang berada di DKI Jakarta. DKI Jakarta sendiri memiliki data harga rumah yang berbeda sesuai dengan kategorinya. Algoritma clustering k-means dalam data mining ini digunakan untuk mengelompokkan data harga rumah berdasarkan fitur-fitur seperti kode pos, daerah, luas tanah, dan harga. Serta untuk mengetahui perbedaan harga kelompok dan menunjukkan harga yang sesuai untuk masing-masing kelompok tersebut. Tujuan membuat laporan tugas akhir ini untuk memberikan informasi kepada masyarakat yang tidak mengetahui apakah data harga rumah DKI Jakarta tersebut termasuk kelompok tinggi, sedang, rendah dan membantu dalam proses pengambilan keputusan dalam properti industri, termasuk pemasaran, pengelompokan harga jual dan pengembangan properti. Berdasarkan hasil penelitian, hasil yang dapat disimpulkan pada data harga rumah DKI Jakarta dapat dikelompokkan menjadi 3 kelompok cluster yaitu Cluster 0 menghasilkan harga rumah sedang sebanyak 193 items, cluster 1 menghasilkan harga rumah rendah sebanyak 1698 items, cluster 2 menghasilkan harga rumah tinggi sebanyak 51 items. Maka dari pengelompokan 3 cluster tersebut menghasilkan nilai optimal dari *Davies Bouldin Index* (DBI) sebesar 0,480.

Kata kunci : Harga Rumah, Clustering, Data Mining, K-Means

1. PENDAHULUAN

Perkembangan sektor properti di Indonesia mengalami peningkatan yang mencolok dalam beberapa tahun terakhir. Faktor kunci dalam sektor properti adalah harga rumah, yang menjadi pertimbangan utama bagi calon pembeli. Daerah Khusus Ibu Kota Jakarta, mengalami pertumbuhan yang pesat dalam pembangunan perumahan. Faktor-faktor seperti pertumbuhan ekonomi, urbanisasi, dan peningkatan infrastruktur di DKI Jakarta telah menyebabkan lonjakan permintaan rumah di wilayah tersebut. Oleh karena itu, memiliki alat pengelompokan yang akurat untuk memproyeksikan data harga rumah di DKI Jakarta menjadi hal yang sangat penting.

Pada kehidupan setiap harinya, manusia memiliki berbagai kebutuhan mendasar, termasuk tempat tinggal. Rumah tidak hanya sebagai tempat beristirahat dan menjaga diri, tetapi memainkan peran penting sebagai pusat kegiatan sehari-hari dan tempat bersosialisasi dengan keluarga serta sanak saudara. Tempat berkumpul adalah rumah yang nyaman. Menyebabkan ketenangan dan ikatan yang kuat di setiap keluarga. Maka, rumah dianggap sebagai keperluan esensial yang diperlukan oleh setiap individu. Pertimbangan utama dalam memilih rumah melibatkan faktor kenyamanan, yang mencakup elemen-elemen seperti lingkungan sekitar, ukuran bangunan, dan harmoni internal keluarga untuk membuat suasana yang nyaman di dalam lingkungan rumah. Selain itu, harga juga merupakan faktor utama yang diperhitungkan dalam proses pemilihan rumah[1].

Penelitian yang dilakukan oleh Briyan Gifari Aji, Dwi Chandra Aditya Sondawa, Muhammad Rifky Gifari, dan Sena Wijayanto pada tahun 2023, berjudul "Penerapan Algoritma K-Means Untuk *Clustering* Harga Rumah Di Bandung". Penerapan yang digunakan dengan metode K-Means *Clustering*. Hasil penelitian menunjukkan bahwa dengan mengatur cluster 2, didapatkan *silhouette score* yang tinggi serta mendekati 1, yakni sebesar 0,8870. Hal ini mengindikasikan bahwa pengelompokan dengan memiliki 2 *cluster*, terdapat interpretasi struktur yang kokoh karena data dalam masing-masing *cluster* sangat rapat dan jelas terpisah dari *cluster* lainnya. [1].

Penelitian yang dilakukan oleh Zyhan Faradilla Daldiri, Muhammad Rafly dan Ionia Veritawati dalam jurnal Journal of Informatics and Advanced Computing (JIAC) pada tahun 2022 yang berjudul "*Clustering* Daftar Harga Rumah di Jakarta Dengan Algoritma K-Means". Penelitian dengan metode K-Means *Clustering*. Dengan menerapkan algoritma K-Means pada *cluster* 5 pada daftar harga rumah di Jakarta, diperoleh hasil klasifikasi, dimulai dari kelas 0 yang memiliki 527 data, kelas 1 dengan 13 data, kelas 2 dengan 108 data, kelas 3 dengan 68 data, dan kelas 4 dengan 294 data. Hasil *silhouette score* mencapai 0,626, yang mengindikasikan bahwa hasil tersebut memiliki interpretasi yang beralasan[2].

Penelitian yang dilakukan oleh Rizky Rahmattullah, Indwiarti dan Aniq Atiqi Rohmawati dalam jurnal e-Proceeding of Engineering yang berjudul "*Clustering* Harga Rumah: Perbandingan Model K-Means dan Gaussian Mixture Model" pada tahun 2023. Penelitian dengan metode *Clustering* K-

Means dan *Gaussian Mixture Model*. Hasil dari kedua proses *clustering* menunjukkan bahwa K-Means memberikan kinerja yang lebih unggul dibandingkan dengan *Gaussian Mixture Model*, dilihat dari dua aspek, yaitu nilai *Silhouette* dan waktu yang dibutuhkan dalam proses *clustering*. K-Means menghasilkan nilai *Silhouette* sebesar 0.63516 dengan waktu proses selama 1.97 detik, sementara *Gaussian Mixture Model* menghasilkan nilai *Silhouette* sebesar 0.62723 dengan waktu proses selama 7.01 detik. Kesimpulan dalam penelitian ini, proses *clustering* menggunakan K-Means menghasilkan hasil yang lebih bagus berbeda dengan *Gaussian Mixture Model*[3].

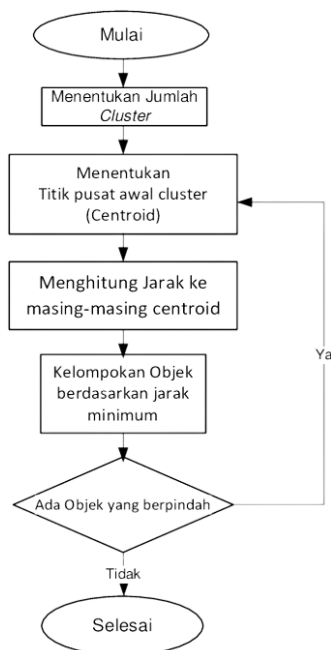
Pada penelitian yang akan dilakukan bertujuan untuk pengelompokan data harga rumah dan mengetahui jumlah *cluster* optimal hasil pengelompokan data harga rumah menggunakan *Davies Bouldin Index* (DBI).

2. TINJAUAN PUSTAKA

2.1. Data Mining

Penambangan data atau data *mining* adalah proses yang menyangkut penyatuan dan penerapan data peristiwa untuk mengidentifikasi pola keterkaitan yang sesuai dalam kumpulan data yang luas[4]. Data *mining* merupakan metode yang menetapkan dan mengakses suatu informasi penting dengan memanfaatkan teknik atau metode khusus yang didasarkan pada pengetahuan atau keterampilan hasil dari penelitian dapat diaplikasikan dalam data mining sebagai pemilihan dalam proses pengambilan keputusan[5].

2.2. K-Means



Gambar 1. Flowchart K-Means

K-Means merupakan metode pengelompokan data non-hirarki yang bertujuan untuk membagi data menjadi berupa *cluster*. Dalam pendekatan ini, data

yang menunjukkan ciri-ciri serupa dikelompokkan bersama dalam 1 *cluster*, sementara data yang menunjukkan ciri-ciri yang berbeda dikelompokkan dalam *cluster* lainnya[6].

Menurut hasil percobaan yang telah dilakukan [7] berikut merupakan tahapannya dalam menerapkan metode k-means *clustering*:

1. Menentukan total *cluster*. Dalam perhitungan, dimulai dengan menetapkan jumlah kelompok *cluster* yaitu *cluster* 2.
2. Menentukan *centroid*. *Centroid* adalah posisi pusat yang dilakukan sebagai awal perhitungan dalam pengelompokan data. Pemilihan awal pusat dilakukan secara acak dari kumpulan data yang digunakan dalam analisis k-means.
3. Menghitung jarak *centroid*. Perhitungan jarak data dan pusat *cluster*. Kedekatan dua objek ditentukan oleh jarak antara keduanya menggunakan *euclidean distance* yang dirumuskan oleh [8] sebagai berikut:

$$d_{\text{Euclidean}}(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (1)$$

Dimana:

- $d(x, y)$: Jarak antara data pada posisi x dan y
 x : posisi data objek
 y : posisi data *centroid*
 i : total atribut pada data

4. Mengelompokkan setiap data pada *cluster* dengan jarak terpendek.
5. Menentukan nilai *centroid* dengan nilai rata-rata *cluster* dari setiap anggota *cluster*.

$$\text{Centroid} = \sum \frac{a_i}{n} \quad (2)$$

Dimana :

- a_i = Nilai anggota masing masing *cluster*
 n = jumlah anggota *cluster*

6. Dilakukan perulangan perhitungan jarak dan menetapkan *cluster* di mana objek tidak mengalami perpindahan objek.

2.3. Clustering

Clustering merupakan proses yang dilakukan untuk mengelompokkan sekumpulan data ke dalam bagian kelompok berdasarkan kesesuaian yang telah dipastikan. Pengelompokan ini disebut sebagai "*cluster*", yang mewakili sekelompok atau kumpulan objek data yang menunjukkan kesamaan dalam *cluster* tertentu dan perbedaan dari objek di *cluster* lain. Objek-objek ini disusun menjadi satu atau lebih *cluster*, memastikan bahwa objek dalam *cluster* memiliki tingkat kesesuaian yang berarti[6].

2.4. Davies Bouldin-Index (DBI)

Davies-Bouldin Index (DBI) adalah salah satu metode evaluasi validitas internal yang digunakan

untuk menilai atau mengevaluasi suatu kluster. DBI diperkenalkan oleh David L. Davies dan Donald W. Bouldin. Validitas internal yang dilakukan DBI adalah seberapa baik kluster dilakukan dengan menghitung kuantitas dan fitur turunan dari kumpulan data.

Tahap-tahap evaluasi kluster dengan menggunakan Davies-Bouldin Index adalah sebagai berikut:

2.4.1. Sum of square within cluster (SSW)

SSW digunakan untuk menjumlahkan kuadrat jarak antara setiap data dalam suatu kluster dan centroid kluster tersebut. rumus SSW adalah sebagai berikut:

$$SSW_i = \sum (X_i - C_i)^2 \quad (3)$$

Keterangan:

m adalah Jumlah data dalam *cluster* ke- i

x adalah data pada *cluster*

c_i = Centroid *cluster* ke- i

$d(X_i, C_i)$ adalah Jarak setiap data ke *centroid* i yang dihitung menggunakan jarak *euclidean*

2.4.2. Sum of square between cluster (SSB)

SSB digunakan untuk menjumlahkan kuadrat jarak antara centroid setiap kluster dan centroid kluster lainnya.

Rumus SSB:

$$SSB = \sum (d(c_i, c_j))^2 \quad (4)$$

Keterangan:

$d(c_i, c_j)$ adalah Jarak *centroid* dengan *centroid* lainnya.

c_i adalah *cluster* satu

c_j adalah lainnya

2.4.3. Pengukuran rasio (Rij)

Pengukuran rasio (R_{ij}) digunakan untuk membandingkan dua kluster berdasarkan nilai kohesi dan separasi. Nilai rasio dihitung menggunakan persamaan sebagai berikut:

$$R_{ij} = SSW_i / SSB_{i,j} \quad (4)$$

Keterangan:

R_{ij} adalah nilai rasio

SSW_i adalah Jumlah kuadrat dalam *cluster* terhadap *centroid* i

$SSB_{i,j}$ adalah Jumlah kuadrat antara *cluster* data i dan j pada *cluster* yang berbeda

2.4.4. Davies-Bouldin Index (DBI)

DBI adalah indeks yang digunakan untuk mengukur kualitas kluster. Nilai rasio yang dihasilkan digunakan untuk mencari nilai *Davies Bouldin Index* (DBI)

Rumus DBI:

$$DBI = \sum (R_{ij}) / K \quad (5)$$

Keterangan:

R_{ij} adalah nilai K = Jumlah *cluster* terbentuk rasio

K adalah jumlah *cluster*

Dari rumus tersebut, k adalah jumlah *cluster* yang digunakan. Semakin kecil nilai DBI yang diperoleh (non-negatif ≥ 0), maka semakin baik kualitas *cluster* diperoleh dari pengelompokan *K-means* tersebut.

2.5. Harga Rumah

Tempat tinggal, yang sering disebut sebagai rumah, adalah suatu wadah bagi manusia yang difungsikan sebagai lokasi untuk tinggal dan melakukan berbagai kegiatan di dalamnya. Rumah merupakan bentuk konstruksi fisik bertujuan sebagai lokasi yang menyediakan perlindungan, tempat untuk tinggal, serta ruang aktivitas bagi individu, kelompok, atau keluarga[1]. Selain itu, Nilai Jual Objek Pajak (NJOP), memiliki variasi di setiap wilayah, dan hal ini menjadi faktor krusial dalam transaksi pembelian rumah. Kehadiran NJOP yang tinggi dapat berfungsi sebagai investasi keuntungan untuk pemilik rumah yang akan menjual rumahnya. Periksa secara teliti kelebihan dan kekurangan rumah, termasuk area plafon atau struktur rumah, serta lakukan pengecekan terhadap potensi kerusakan dan perluasan yang mungkin dibutuhkan. Selanjutnya, bandingkan harga rumah dengan nilai pasar dan pertimbangkan lokasi hunian dengan seksama, karena lebih strategis dan berdekatan dengan fasilitas, juga dapat diakses dengan mudah, maka kualitas rumah dan akibatnya harganya akan meningkat[9].

2.6. RapidMiner

RapidMiner merupakan suatu perangkat lunak atau aplikasi dengan AGPL (GNU Affero General Public License) yang berkarakter *open source*. Aplikasi ini dibesarkan oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di Unit Kecerdasan Buatan dari *University of Dortmund*. Fungsi utama *RapidMiner* adalah untuk melakukan pengolahan Data Mining. Aplikasi ini juga menangani area seperti *text mining* dan analisis data. Pengembangan *RapidMiner* bertujuan memberikan solusi bagi analisis di berbagai bidang, termasuk Data Mining dengan dilakukan berbagai metode deskriptif dan prediktif. Pada pengaplikasiannya, *RapidMiner* menyediakan sekitar 500 operator Data Mining, yang mencakup *operator input*, *operator output*, *preprocessing* data, dan visualisasinya. *RapidMiner* merupakan bahasa pemrograman Java, memungkinkannya beroperasi pada semua jenis sistem operasi[10].

3. METODE PERANCANGAN

3.1. Sumber Data

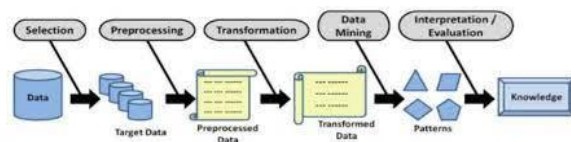
Data yang akan digunakan pada penelitian ini merupakan data sekunder berupa kumpulan data harga rumah DKI Jakarta yang diperoleh dari website <https://www.rumah123.com/jual/dki-jakarta/rumah/>. Dataset ini menghasilkan atribut mengenai deskripsi, kode pos, daerah, kamar, kamar mandi, garasi, luas tanah, luas bangunan, dan harga. Namun atribut yang

akan digunakan pada penelitian ini terdiri dari kode pos, daerah, luas tanah, dan harga dengan jumlah data dalam dataset mencapai 1942 data.

3.2. Teknik Pengumpulan Data

Pengumpulan data yang akan dilakukan pada penelitian ini melalui observasi. Namun untuk penelitian observasi tidak dilakukan secara langsung ke dalam tempat atau perusahaan tertentu, mengingat data yang digunakan diperoleh dari internet. Oleh karena itu, hanya dilakukan peninjauan data yang bertujuan untuk mendapatkan dataset yang sesuai dengan topik penelitian ini. Lalu dilakukan teknik studi pustaka atau studi literatur untuk mencari referensi penelitian berupa jurnal dengan topik pembahasan dan metode yang sesuai dengan penelitian ini.

3.3. Tahapan Perancangan



Gambar 2. Proses KDD

Penelitian ini melibatkan langkah-langkah pada data mining digunakan proses *Knowledge Discovery in Database* (KDD), yang merujuk pada proses mencari informasi tersembunyi pada kumpulan data luas. KDD sering digunakan untuk menjelaskan proses mencari informasi yang masih belum diketahui pada suatu basis data yang luas.

Adapun prosesnya sebagai berikut:

1. *Data Selection*, yaitu langkah pengambilan, pemilihan, dan pengecekan informasi.
2. *Data Preprocessing*, merupakan proses pembersihan data dengan menghapus data yang tidak perlu, data hilang dan missing.
3. *Data Transformation*, proses *transformation* data ke dalam format sesuai untuk proses data mining.
4. *Data Mining*, merupakan proses pengolahan data berdasarkan alur data mining.
5. *Data Interpretation/Evaluation*, digunakan untuk mengevaluasi hasil dari penelitian.

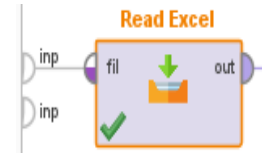
4. HASIL DAN PEMBAHASAN

4.1. Implementasi K-Means Clustering Untuk Pengelompokan Data Harga Rumah.

Langkah-langkah Implementasi K-Means Clustering Untuk Pengelompokan Data Harga Rumah:

4.1.1. Data Selection

Dataset yang akan digunakan untuk pemodelan K-Means Clustering dengan digunakannya data harga rumah DKI Jakarta memiliki atribut deskripsi, kode pos, daerah, kamar, kamar mandi, garasi, luas tanah, luas bangunan, dan harga.



Gambar 3. Operator read excel

Pada gambar di atas, langkah pertama yang harus dilakukan yaitu *read excel* digunakan untuk membaca data dari *file excel* ke dalam *rapidminer*.

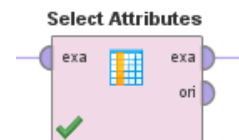
Gambar di atas menghasilkan data *excel* yang telah di *import* dari operator *read excel*.

4.1.2. Data Preprocessing

Pada proses *Preprocessing* dari data harga rumah DKI Jakarta tidak memiliki data *missing*. Oleh karena itu, tidak dilakukannya penghapusan data yang kosong. Selanjutnya dilakukan proses *select attributes*.

	A	B	C	D	E	F	G	H	I
1	deskripsi	harga (Rp...)	kode pos	daerah	kamar	kamarMa...	garasi	luas tanah	luas ban...
2	Rumah M...	37000.000	12510.000	Jakarta S...	4	4	8	1600.000	900.000
3	Rumah b...	6900.000	12440.000	Jakarta S...	4	3	2	263.000	380.000
4	Dijual Ce...	8000.000	12510.000	Jakarta S...	4	3	2	210.000	350.000
5	Rumah P...	4990.000	11610.000	Jakarta B...		3	3	144.000	200.000
6	Rumah m...	990.000	13910.000	Jakarta TL...	2	2	1	48.000	73.000
7	Rumah 2...	1360.000	10510.000	Jakarta P...	2	2	1	40.000	77.000
8	Rumah el...	1500.000	12620.000	Jakarta S...	4	4	2	71.000	130.000
9	Rumah ta...	2550.000	11470.000	Jakarta B...	3	3	2	60.000	200.000
10	Rumah b...	2250.000	11470.000	Jakarta B...	3	3	1	56.000	150.000
11	Temurah...	14500.000	14460.000	Jakarta Ut...		4	3	320.000	400.000
12	Rumah D...	3900.000	15147.000	Jakarta B...	2	3	2	144.000	200.000
13	Rumah Sl...	4250.000	12810.000	Jakarta S...	4	4	2	226.000	325.000
14	Rumah P...	11500.000	11610.000	Jakarta B...	5	3	3	403.000	315.000

Gambar 4. Import read excel

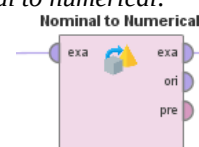


Gambar 5. Operator select attributes

Dari analisis data harga rumah DKI Jakarta atribut yang akan digunakan hanya kode pos, daerah, luas tanah, dan harga menggunakan operator *select attributes* pada gambar 4.

4.1.3. Data Transformation

Pada Tahapan ini, data harga rumah DKI Jakarta mencantumkan data *transformation* dengan operator *nominal to numerical*.

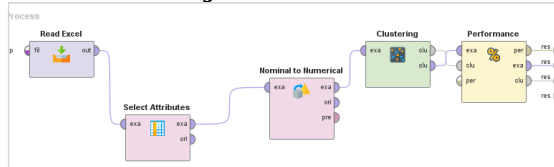


Gambar 6. Operator Nominal to numerical

Dikarenakan gambar di atas mengubah tipe atribut non-numerik yang dipilih menjadi tipe numerik

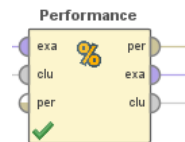
pada salah satu atribut daerah yang sesuai pada alur data mining.

4.1.4. Data Mining



Gambar 7. Data Mining

Teknik yang dilakukan pada penelitian ini dengan data mining yaitu k-means clustering. Untuk membantu mengelompokkan data harga rumah DKI Jakarta. Penelitian ini menggunakan aplikasi RapidMiner versi 10. Gambar di atas adalah operator utama dalam proses pemodelan untuk mengevaluasi hasil pengelompokan dengan menguji nilai K hingga 10 cluster.



Gambar 8. Performance

Operator Performance untuk mengukur Cluster Distance Performance dari performa operator K-Means dan memberikan daftar nilai kriteria kinerja berdasarkan centroid dari setiap cluster. 2 kinerja mencakup nilai rata-rata jarak pada cluster dan metode Davies Bouldin.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 30814876.033
Avg. within centroid distance_cluster_0: 78721623.947
Avg. within centroid distance_cluster_1: 20796439.591
Avg. within centroid distance_cluster_2: 183075713.875
Davies Bouldin: 0.480
```

Gambar 9. Hasil performance vector

Hasil performance di atas merupakan nilai rata-rata dalam jarak cluster dari 3 cluster menghasilkan nilai Davies Bouldin Index sebesar 0,480. Semakin mendekati angka 0 pada Davies Bouldin Index menunjukkan kualitas yang semakin baik.

4.1.5. Interpretation / Evaluation

Pada tahapan ini menunjukkan hasil pengelompokan data harga rumah DKI Jakarta menggunakan algoritma k-means.

Cluster Model

```
Cluster 0: 193 items
Cluster 1: 1698 items
Cluster 2: 51 items
Total number of items: 1942
```

Gambar 10. Cluster model

Pada penelitian ini sesuai gambar di atas pengelompokan dengan algoritma k-means dibagi

menjadi 3 cluster pada cluster 0 sebanyak 193 items, cluster 1 sebanyak 1698 items dan cluster 2 sebanyak 51 items dari total data 1942 items.



Gambar 11. Visualisasi harga rumah dan daerah

Visualisasi di atas menghasilkan 3 cluster dari pengelompokan 5 daerah diantaranya daerah 0 yaitu jakarta selatan, daerah 1 yaitu jakarta barat, daerah 2 yaitu jakarta timur, daerah 3 yaitu jakarta pusat, dan daerah 4 yaitu jakarta utara.

Tabel .2 Pengelompokan harga rumah dan luas tanah

Cluster	Harga	Luas Tanah	Keterangan
0	18500 - 50000	200 – 2543	Harga Rumah Sedang
1	235 - 18000	19 – 3600	Harga Rumah Rendah
2	53900 - 99000	474 – 8200	Harga Rumah Tinggi

Pada tabel 1, Diperoleh 3 cluster yaitu cluster 0 dari harga 18500 sampai 50000 dengan luas tanah 200 sampai 2543 meter persegi menghasilkan kelompok harga rumah sedang. Cluster 1 dari harga 235 sampai 18000 dengan luas tanah 19 sampai 3600 meter persegi menghasilkan kelompok harga rumah rendah dan cluster 2 dari harga 53900 sampai 99000 dengan luas tanah 474 sampai 8200 meter persegi menghasilkan kelompok harga rumah tinggi.

4.2. Nilai DBI Optimal Hasil Pengelompokan Data Harga Rumah

Davies Bouldin Index (DBI) digunakan untuk mengevaluasi kinerja proses pengelompokan, menghasilkan parameter numerik. Semakin kecil nilai yang diberikan, semakin baik kinerja cluster. Proses percobaan ini dilakukan dengan mencari cluster optimal mulai dari 2 hingga 10 cluster.

Tabel 3 Hasil DBI

Jumlah Cluster	Davies Bouldin Index (DBI)
2	0.482
3	0.480
4	0.513
5	0.539
6	0.530
7	0.566
8	0.593

Jumlah Cluster	Davies Bouldin Index (DBI)
9	0.580
10	0.582

Hasil optimal ditemukan menggunakan model *cluster* 3 yang memiliki akurasi lebih kecil, menghasilkan *Davies Bouldin Index* sebesar 0,480. Dengan *k-means clustering* efektif untuk mengelompokan data harga rumah DKI Jakarta, dan hasil terbaik dicapai dengan 3 *cluster*, dengan nilai *Davies Bouldin Index* terbaik sebesar 0,480 pada *cluster* 3.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil dari penelitian dan pembahasan pada data harga rumah Jakarta menggunakan 1942 data tahun 2023. Pengelompokan data diolah melalui *tools RapidMiner* dengan menggunakan algoritma *K-Means clustering*. Maka akan memperoleh kesimpulan diantaranya dapat dikelompokkan menjadi 3 *cluster* yaitu *cluster* tinggi, sedang dan rendah. *Cluster* 0 (sedang), sebanyak 193 *items* dari harga 18500 sampai 50000 dengan luas tanah 200 sampai 2543 meter persegi menghasilkan banyaknya daerah pada jakarta selatan sebanyak 134 daerah. *Cluster* 1 (rendah), sebanyak 1698 *items* dari harga 235 sampai 18000 dengan luas tanah 19 sampai 3600 meter persegi menghasilkan banyaknya daerah pada jakarta selatan sebanyak 523 daerah. *Cluster* 2 (Tinggi), sebanyak 51 *items* dari harga 53900 sampai 99000 dengan luas tanah 474 sampai 8200 meter persegi menghasilkan banyaknya daerah pada jakarta selatan sebanyak 33 daerah. Hasil pencarian nilai optimal *Davies Bouldin Index* terbaik dari pengelompokan harga rumah dengan algoritma *k-means clustering* menggunakan 3 *cluster* menghasilkan nilai sebesar 0,480. Saran pada penelitian ini agar dapat memperoleh hasil yang lebih bagus dari penelitian ini sebaiknya dilakukan penelitian lebih lanjut dengan menerapkan metode lainnya seperti Metode *Fuzzy C-Means*, *K-Medoids*, dan *X-Means* sehingga hasil yang didapatkan bisa lebih baik dan dapat digunakan sebagai bahan perbandingan pada hasil yang didapatkan. Penelitian ini menghasilkan *cluster* pada data harga rumah DKI Jakarta, dengan adanya pengelompokan ini diharapkan dapat mempermudah untuk masyarakat khususnya pembeli rumah di DKI Jakarta agar lebih baik dalam pemilihan rumah sesuai dengan daerah yang diinginkan dan harga yang sesuai.

DAFTAR PUSTAKA

- [1] B. G. Aji, D. C. A. Sondawa, M. R. Gifari, and S. Wijayanto, "Penerapan Algoritma K-Means Untuk Clustering Harga Rumah Di Bandung," *J. Ilm. Inform. Glob.*, vol. 14, no. 2, pp. 17–23, 2023.
- [2] Z. Faradilla Daldiri -, M. Rafly -, and I. Veritawati -, "Clustering Daftar Harga Rumah di Jakarta Dengan Algoritma K-Means," *J. Informatics Adv. Comput.*, vol. 3, no. 2, 2022, [Online]. Available: <https://www.kaggle.com/datasets/wisnuanggara/daf>
- [3] R. Rahmattullah, Indwiarti, and A. A. Rohmawati, "Clustering Harga Rumah: Perbandingan Model K-Means dan Gaussian Mixture Model," *e-Proceeding Eng.*, vol. 10, no. 3, pp. 3441–3449, 2023.
- [4] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [5] T. Hardiani, "Analisis Clustering Kasus Covid 19 di Indonesia Menggunakan Algoritma K-Means," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 2, pp. 156–165, 2022, doi: 10.23887/janapati.v11i2.45376.
- [6] S. Aulia, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja)," *Djtechno J. Teknol. Inf.*, vol. 1, no. 1, pp. 1–5, 2021, doi: 10.46576/djtechno.v1i1.964.
- [7] L. P. Refialy, H. Maitimu, and M. S. Pesulima, "Perbaikan Kinerja Clustering K-Means pada Data Ekonomi Nelayan dengan Perhitungan Sum of Square Error (SSE) dan Optimasi nilai K cluster," *Techno.Com*, vol. 20, no. 2, pp. 321–329, 2021, doi: 10.33633/tc.v20i2.4572.
- [8] A. Rohmah, F. Sembiring, and A. Erfina, "Implementasi Algoritma K-Means Clustering Analysis untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: SMK Yaspim Gegerbitung)," *Semin. Nas. Sist. Inf. dan Manaj. Inform.*, pp. 290–298, 2021, [Online]. Available: <https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/32>
- [9] M. L. Mu'tashim, T. Muhayat, S. A. Damayanti, H. N. Zaki, and R. Wirawan, "Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression," *Inform. J. Ilmu Komput.*, vol. 17, no. 3, p. 238, 2021, doi: 10.52958/iftk.v17i3.3635.
- [10] Dahlia and Andri, "Implementasi Data Mining untuk Prediksi Persediaan Obat pada Puskesmas Kertapati menggunakan Regresi Linier Berganda," *J. Sist. dan Inform.*, vol. 15, no. 2, pp. 95–103, 2020, doi: 10.30864/jsi.v15i2.331.