

Penerapan Algoritma K-Means Clustering Untuk Harga Rumah di Jakarta Selatan

Nuraeni Septiani

Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) IKMI Cirebon

Saeful Anwar

Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) IKMI Cirebon

Ruli Herdiana

Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) IKMI Cirebon

Alamat: Cirebon, Jawa Barat, Indonesia 45131

Korespondensi penulis: nuraeniseptiani2@gmail.com

Abstract. *The price of a house is the value or cost assigned to a residential property, usually expressed in a certain currency. Home prices are determined by various factors, such as location, size, condition, facilities owned, as well as market factors such as demand and competition. House prices are a useful tool for analyzing and understanding the housing market in a given area by displaying information about house prices in that area in an organized and easy-to-read format. The k-means clustering algorithm is used to classify house price data based on features such as location, size, type of house, and so on. The purpose of using the k-means clustering algorithm is to find out the price differences between groups and determine the appropriate price for each group. The results of this analysis can be used to assist in the decision-making process in industrial property, including marketing, determining the selling price, and property development. The data collection technique that the researcher chose used secondary data techniques, which are sources of research data that can produce a number of collected data. during research conducted quickly and informally through the use of intermediary media. The research source that the researchers chose came from Kaggle, with a total of 1010 data on house prices. Based on the research results, it can be concluded that house prices in South Jakarta can be grouped into 10 clusters according to the best dbi value that is 0.129.*

Keywords: *K-Means Clustering, Secondary Data, House Prices, Hous, Dbi Value.*

Abstrak. Harga rumah adalah nilai atau biaya yang ditetapkan untuk sebuah properti perumahan, biasanya dinyatakan dalam mata uang tertentu. Harga rumah ditentukan oleh berbagai faktor, seperti lokasi, ukuran, kondisi, fasilitas yang dimiliki, serta faktor pasar seperti permintaan dan persaingan. Harga rumah merupakan alat yang berguna untuk menganalisis dan memahami pasar perumahan di suatu wilayah tertentu dengan menampilkan informasi tentang harga rumah di wilayah tersebut dalam format yang terorganisir dan mudah dibaca. Algoritma clustering k-means ini digunakan untuk mengelompokkan data harga rumah berdasarkan fitur-fitur seperti lokasi, ukuran, jenis rumah, dan lain-lain. Tujuan dari penggunaan algoritma clustering k-means ini adalah untuk mengetahui perbedaan harga antar kelompok dan menentukan harga yang sesuai untuk masing-masing kelompok. Hasil analisis ini dapat digunakan untuk membantu

dalam proses pengambilan keputusan dalam properti industri, termasuk pemasaran, penentuan harga jual, dan pengembangan properti.. Teknik pengumpulan data yang peneliti pilih menggunakan teknik data sekunder, yang merupakan sumber data penelitian yang dapat menghasilkan sejumlah data yang terkumpul selama penelitian yang dilakukan secara cepat dan informal melalui penggunaan media perantara. Sumber penelitian yang peneliti pilih berasal dari Kaggle total keseluruhan data harga rumah tersebut yaitu 1010 data. Berdasarkan hasil penelitian, maka dapat disimpulkan harga rumah di Jakarta Selatan dapat dikelompokkan menjadi 10 cluster sesuai dengan nilai dbi terbaik yaitu 0.129.

Kata kunci: Clustering K-Means, Data Sekunder, Harga Rumah, Nilai Dbi.

LATAR BELAKANG

Harga rumah adalah nilai atau biaya yang ditetapkan untuk sebuah properti perumahan, biasanya dinyatakan dalam mata uang tertentu. Harga rumah ditentukan oleh berbagai faktor, seperti lokasi, ukuran, kondisi, fasilitas yang dimiliki, serta faktor pasar seperti permintaan dan persaingan. Harga rumah merupakan alat yang berguna untuk menganalisis dan memahami pasar perumahan di suatu wilayah tertentu dengan menampilkan informasi tentang harga rumah di wilayah tersebut dalam format yang terorganisir dan mudah dibaca. Terlepas dari faktor penting lainnya, lokasi memiliki efek peredam yang signifikan terhadap keinginan konsumen untuk membeli. Lokasi rumah yang dituju konsumen biasanya berada di kawasan strategis yang dekat dengan tempat kerja, tempat perbelanjaan, dan tempat ibadah. Bagi kelompok yang memiliki anak, pemisahan antara tempat tinggal dan tempat pendidikannya juga menjadi hal yang harus diperhatikan.

Konsep dari penerapan algoritma K-Means Clustering untuk analisis harga rumah di Jakarta Selatan adalah mengelompokkan data rumah berdasarkan kesamaan fitur yang dimiliki, seperti lokasi, ukuran, dan fasilitas, dengan tujuan untuk menemukan pola atau struktur dalam harga rumah di Jakarta Selatan. Setelah data rumah terkelompok, maka dapat menganalisis perbedaan harga antar kelompok dan menentukan harga yang sesuai untuk tiap kelompok. Hal ini dapat membantu dalam proses mengambil keputusan dalam industri properti seperti pemasaran, penentuan harga jual, dan pengembangan properti.

Adapun pokok masalah dari penerapan algoritma k-means clustering harga rumah di Jakarta selatan yaitu, sulitnya memahami pasar perumahan di Jakarta Selatan. Pasar perumahan di Jakarta maupun di kota-kota besar lainnya cukup kompleks dan dinamis, dengan banyak faktor yang mempengaruhi harga rumah, seperti lokasi, tipe, umur, kondisi, dan infrastruktur. Kompleksitas ini dapat mempersulit identifikasi pola dan tren

di pasar, terutama saat menganalisis data mentah. Maka dari itu peneliti akan mengelompokkan dengan algoritma clustering k-means supaya mudah memahami pasar perumahan di Jakarta Selatan dan menentukan cluster optimum dan nilai dbi dari hasil mengelompokkan Harga Rumah menggunakan algoritma k-means clustering. Clustering adalah satu-satunya teknik yang digunakan dalam fitur penambangan data mining.

Tujuan dari penelitian penerapan algoritma k-means clustering harga rumah di Jakarta selatan adalah untuk mengelompokkan harga rumah menggunakan algoritma k-means clustering dan menentukan cluster optimum dan nilai dbi dari hasil mengelompokkan Harga Rumah menggunakan algoritma k-means clustering.

Adapun tujuan dari batasan masalah tersebut yaitu untuk membatasi ruang lingkup penelitian, guna penelitian ini terfokus pada Dataset yang digunakan yaitu data Harga Rumah tahun 2020 sebanyak 1010 record dan Rapidminer versi 10.0.0.

KAJIAN TEORITIS

A. Clustering

Clustering adalah satu-satunya teknik yang digunakan dalam fitur penambangan data mining. Algoritma clustering melibatkan transfer beberapa data ke inti kumpulan data yang relevan (Studi & Informatika, 2022) Proses memisahkan satu set objek data ke dalam kantong terpisah dikenal sebagai clustering (Apulino Iman Seno Aji et al., 2021)

B. Data Mining

Data Mining adalah proses mengekstraksi dan mengidentifikasi informasi yang dapat digunakan dan mengumpulkan pengetahuan dari kumpulan data besar menggunakan pendekatan statistik, matematika, kecerdasan buatan, dan Machine Learning. (Ginting et al., 2019) Ada subbidang lain dari Data Mining, seperti knowledge discover ataupun pattern recognition. (Bastian et al., n.d.)

C. K-Means

K-Means merupakan satu-satunya teknik data mining yang dapat mengelompokkan data ke dalam satu cluster atau lebih, Data dengan karakteristik yang mirip dapat ditambahkan ke cluster yang bersangkutan, dan data dengan karakteristik yang berbeda dapat ditambahkan ke cluster lain yang serupa dengan cluster yang bersangkutan. tersebut yang awalnya mereka kelompokkan (Dinata et al., 2020) Terdapat dua proses dalam melaksanakan Algoritma K-Means diantaranya

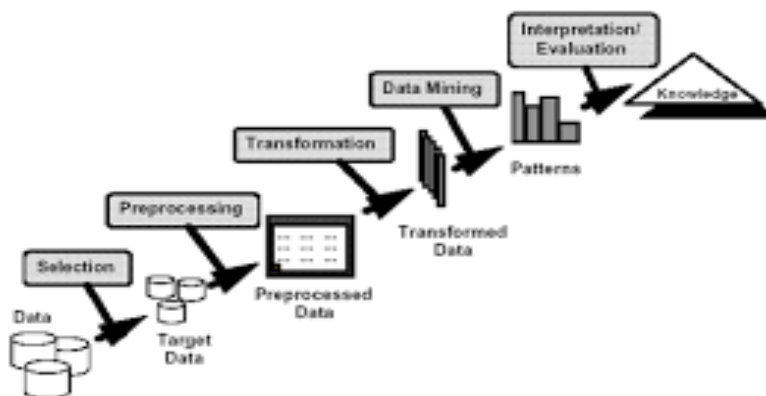
mengidentifikasi lokasi pusat setiap cluster dan mengidentifikasi setiap anggota cluster(Sagala, 2021)

D. RapidMiner

RapidMiner adalah perangkat lunak yang mudah diakses dan tersedia secara konsisten (open source)(Mohede et al., 2020) RapidMiner adalah salah satu software untuk pengolahan data mining(Ramdhan et al., 2022)

METODE PENELITIAN

Tahapan perancangan yang digunakan dalam penelitian ini menggunakan rancangan Knowledge Discovery In Databases (KDD). Knowledge Discovery in Database (KDD) adalah proses untuk memahami pengetahuan dalam database. KDD didefinisikan secara rinci sebagai prosedur untuk mengumpulkan pengetahuan dan informasi potensial dari kumpulan data yang besar. Pengetahuan dan informasi yang diperoleh KDD bersifat sah, baru, mudah dimengerti, dan bermanfaat. Adapun KDD yaitu nama lain dari Data Mining (Ginting et al., 2019)



Gambar 1 Tahapan Perancangan

Tugas data mining meliputi selection data dari sumber data ke data target, preprocessing data untuk meningkatkan kualitas data, transformation data, data mining, dan tugas interpretation dan evaluation yang menghasilkan pengetahuan baru yang dimaksudkan untuk memberikan kontribusi yang lebih bermanfaat. (Ramdhan et al., 2022)

1. Data Selection

Data Selection merupakan Pemilihan (seleksi) Data adalah proses pemilihan data dari kumpulan data operasional, dan harus diselesaikan Sebelum dimulainya proses pengumpulan informasi KDD, data dari proses seleksi digunakan untuk data mining terdapat dalam satu dokumen dan berasal dari dasar data operasional.(Fatmawati & Windarto, 2018)

2. Data Preprocessing

Data Preprocessing adalah langkah pertama dalam proses data mining. Sebelum melanjutkan ke tahap pemrosesan. Data akan dirilis lebih cepat. Preprocessing data, juga dikenal sebagai persiapan data, sering dilakukan dengan menghilangkan data yang tidak relevan.

3. Data Transformation

Data Transformation digunakan untuk mengubah data menjadi format yang sesuai untuk proses data mining. Normalization, pemilihan attribute, dan discretization adalah beberapa teknik untuk transformasi data

4. Data Mining

Proses di mana berbagai teknik digunakan untuk mengidentifikasi pola-pola potensial menghasilkan data yang berguna. Data Mining biasanya memiliki komponen yang berasal dari bidang pengetahuan, seperti pembelajaran (artificial intelligent) machine learning, statistik dan database (Purba et al., 2019)

5. Interpretation / evaluation

Langkah ini mengurangi kebutuhan untuk menentukan apakah data atau informasi yang disajikan konsisten dengan fakta atau hipotesis sebelumnya.

Teknik Pengumpulan Data

Teknik Pengumpulan data yaitu dengan melakukan teknik online research, adapun teknik online research adalah melakukan penelusuran data melalui internet dengan mengakses situs web kaggle. Data set yang diambil yaitu data Harga Rumah di Jakarta Selatan pada tahun 2020. Jumlah record data dari dataset ini sebanyak 1010 data.

Jenis data yang digunakan dalam analisis ini adalah data sekunder. Data Sekunder adalah informasi yang sudah dikumpulkan oleh peneliti dari rangkuman yang sudah ada. Data sekunder merupakan informasi dan data yang diperoleh dari berbagai buku, publikasi, jurnal ilmiah, dan sumber online.(Al-rizki et al., 2020) Sumber data sekunder yang digunakan dalam penelitian ini didasarkan pada data yang tersedia dari website kaggle.

HASIL DAN PEMBAHASAN

A. Penerapan Algoritma Clustering K-Means Untuk Harga Rumah

1. Pengumpulan Data

Sumber Data yang digunakan pada penelitian ini adalah dataset Harga Rumah di Jakarta Selatan. Sumber Data yang digunakan pada penelitian ini adalah dataset Harga Rumah di Jakarta Selatan, kemudian data-data tersebut dipelajari dan dikelompokkan (R. M. Sari et al., 2020)

Tabel 1 Pengumpulan Data

No	Nama Rumah	Harga	LB	LT	KT	KM	GRS
1	Rumah Murah Hook Tebet Timur, Tebet, Jakarta Selatan	3800000000	220	220	3	3	0
2	Rumah Modern di Tebet dekat Stasiun, Tebet, Jakarta Selatan	4600000000	180	137	4	3	2
3	Rumah Mewah 2 Lantai Hanya 3 Menit Ke Tebet, Tebet, Jakarta Selatan	3000000000	267	250	4	4	4
4	Rumah Baru Tebet, Tebet, Jakarta Selatan	4300000000	40	25	2	2	0
5	Rumah Bagus Tebet komp Gudang Peluru lt 350m, Tebet, Jakarta Selatan	9000000000	400	355	6	5	3
6	Rumah Mewah Modern Murah 3 lantai di Tebet Jakarta Selatan, Tebet, Jakarta Selatan	4970000000	300	154	5	3	3
7	Rumah lama di Tebet, dekat MT Haryono dan tol dalam kota, jalan 1 mobil hanya 30M dr jln besar, Tebet, Jakarta Selatan	2600000000	120	150	3	2	1
8	Rumah Bagus Keren Jalan Lebar di Area & Kawasan Terbaik Tebet, Tebet Jakarta Selatan	10500000000	350	247	4	4	0
9	Minimalis Baru Jalan 1 Mobil Akses Mudah Dekat ke Jalan Lebar, Tebet, Jakarta Selatan	3250000000	125	90	3	3	0
10	...	...	...	...	...	...	...

1006	Rumah Strategis Akses Jalan 2mobil Di Menteng Dalam Tebet Jak Sel	9000000000	450	550	10	10	3
1007	Tebet Rumah Siap Huni Jln 2 Mbl Nyaman	4000000000	160	140	4	3	2
1008	Di Kebun Baru Rumah Terawat, Area Strategis	4000000000	139	230	4	4	1
1009	Dijual Cepat Rumah Komp Depkeu Dr Soepomo Tebet Jakarta Selatan	19000000000	360	606	7	4	0
1010	Dijual Rumah Kokoh Di Gudang Peluru	10500000000	420	430	7	4	2

2. Data Selection

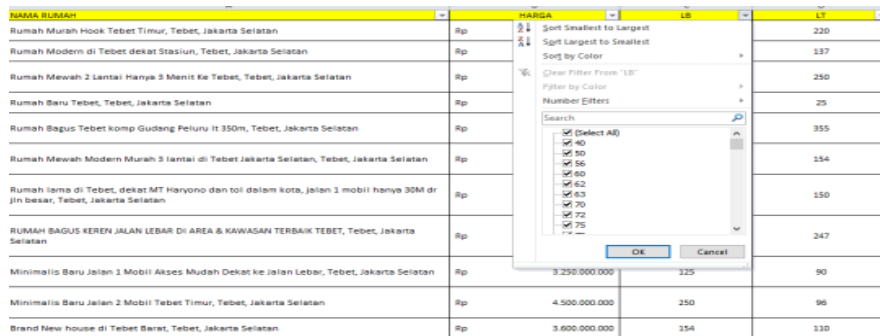
Dataset harga rumah di Jakarta Selatan lebih sering diperbarui dan beberapa atribut tabel tersedia untuk analisis. Ada empat atribut yang digunakan, antara lain Nama Rumah, Harga, LB (Luas Bangunan), dan LT (Luas Tanah).

NAMA RUMAH	HARGA	LB	LT
Rumah Murah Hook Tebet Timur, Tebet, Jakarta Selatan	Rp 3.800.000.000	220	220
Rumah Modern di Tebet dekat Stasiun, Tebet, Jakarta Selatan	Rp 4.600.000.000	180	197
Rumah Mewah 2 Lantai Hanya 3 Menit Ke Tebet, Tebet, Jakarta Selatan	Rp 5.000.000.000	267	250
Rumah Baru Tebet, Tebet, Jakarta Selatan	Rp 430.000.000	40	25
Rumah Bagus Tebet komp Gudang Peluru It 350m, Tebet, Jakarta Selatan	Rp 9.000.000.000	400	355
Rumah Mewah Modern Murah 3 lantai di Tebet Jakarta Selatan, Tebet, Jakarta Selatan	Rp 4.970.000.000	300	154
Rumah lama di Tebet, dekat MT Hanyono dan tol dalam kota, jalan 1 mobil hanya 30M dr Jln besar, Tebet, Jakarta Selatan	Rp 2.600.000.000	120	150
RUMAH BAGUS KEREN JALAN LEBAR DI AREA & KAWASAN TERBAIK TEBET, Tebet, Jakarta Selatan	Rp 10.500.000.000	350	247
Minimalis Baru Jalan 1 Mobil Akses Mudah Dekat ke Jalan Lebar, Tebet, Jakarta Selatan	Rp 3.250.000.000	125	90
Minimalis Baru Jalan 2 Mobil Tebet Timur, Tebet, Jakarta Selatan	Rp 4.500.000.000	250	96
Brand New house di Tebet Barat, Tebet, Jakarta Selatan	Rp 3.600.000.000	154	110

Gambar 2 Data Selection

3. Data Preprocessing

Setelah data dimasukkan dan dipilih sesuai dengan atribut yang akan digunakan, pre-processing data dilakukan untuk memastikan tidak ada entri duplikat, tidak ada nilai yang hilang, dan untuk menghilangkan potensi kesalahan dari dataset yang baru dalam format excel. Pembersihan atau penyiapan data akan dilakukan saat ini agar data mining dapat dilakukan dengan data yang relevan.

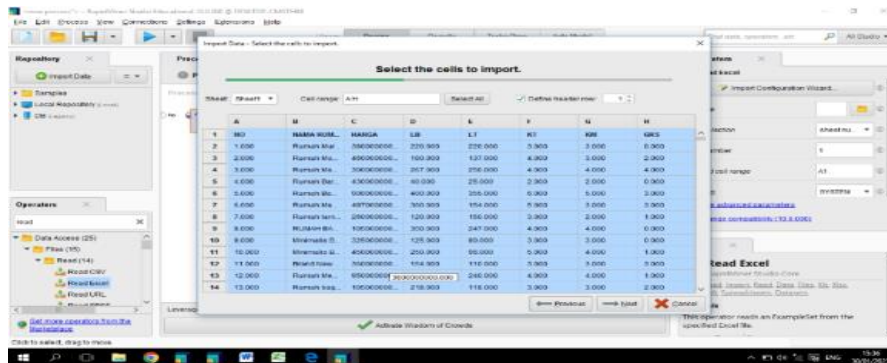


NAMA RUMAH	HARGA	LB	LT
Rumah Murah Hook Tebet Timur, Tebet, Jakarta Selatan	Rp		220
Rumah Modern di Tebet dekat Stasiun, Tebet, Jakarta Selatan	Rp		137
Rumah Mewah 2 Lantai Hanya 5 Menit Ke Tebet, Tebet, Jakarta Selatan	Rp		250
Rumah Baru Tebet, Tebet, Jakarta Selatan	Rp		25
Rumah Bagus Tebet komp Gudang Peluru It 350m, Tebet, Jakarta Selatan	Rp		355
Rumah Mewah Modern Murah 3 lantai di Tebet Jakarta Selatan, Tebet, Jakarta Selatan	Rp		154
Rumah lama di Tebet, dekat MT Haryono dan tol dalam kota, jalan 1 mobil hanya 30M dr Jln besar, Tebet, Jakarta Selatan	Rp		150
RUMAH BAGUS KEREN JALAN LEBAR DI AREA & KAWASAN TERBAIK TEBET, Tebet, Jakarta Selatan	Rp		247
Minimalis Baru Jalan 1 Mobil Akses Mudah Dekat ke Jalan Lebar, Tebet, Jakarta Selatan	Rp	3.250.000.000	125
Minimalis Baru Jalan 2 Mobil Tebet Timur, Tebet, Jakarta Selatan	Rp	4.500.000.000	250
Brand New house di Tebet Barat, Tebet, Jakarta Selatan	Rp	3.600.000.000	154
			110

Gambar 3 Data Preprocessing

4. Data Transformation

Pemutakhiran data dalam format yang sesuai untuk proses data mining sedang dilakukan saat ini. Salah satu teknik untuk transformation data adalah pemberian tag atribut.

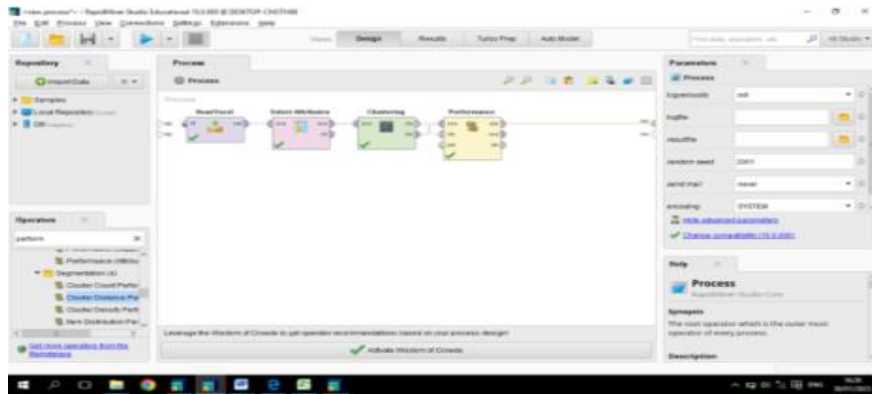


NO	NAMA RUMAH	HARGA	LB	LT	RT	KM	GES
1	1.000	Rumah Mewah	550000000	220.000	220.000	5.000	3.000
2	2.000	Rumah Mewah	480000000	160.000	137.000	4.000	2.000
3	3.000	Rumah Mewah	300000000	267.000	270.000	4.000	4.000
4	4.000	Rumah Baru	430000000	40.000	25.000	2.000	0.000
5	5.000	Rumah Mewah	580000000	400.000	350.000	5.000	5.000
6	6.000	Rumah Mewah	497000000	300.000	160.000	5.000	3.000
7	7.000	Rumah Mewah	280000000	120.000	150.000	3.000	2.000
8	8.000	Rumah Mewah	100000000	300.000	247.000	4.000	0.000
9	9.000	Minimalis Baru	325000000	125.000	85.000	3.000	3.000
10	10.000	Minimalis Baru	450000000	250.000	50.000	5.000	1.000
11	11.000	Brand New	360000000	154.000	110.000	5.000	3.000
12	12.000	Rumah Mewah	550000000	220.000	240.000	4.000	3.000
13	13.000	Rumah Mewah	480000000	160.000	137.000	4.000	2.000

Gambar 4 Data Transformation

5. Data Mining

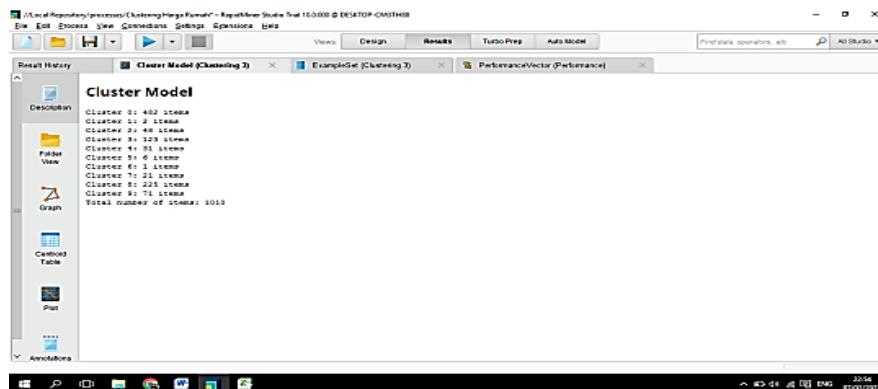
Teknik data mining yang digunakan dalam penelitian ini adalah Clustering K-Means. Untuk membantu prosedur pengolahan data. Penelitian ini menggunakan perangkat lunak RapidMiner versi 10. Berdasarkan gambar 5, algoritma k-means menggunakan operator read excel yang digunakan untuk memanggil dataset atau membuka file yang berekstensi xls (data dalam bentuk excel), setelah itu menggunakan operator select atribut untuk memilih fitur dalam satu layer, kemudian menggunakan clustering k-means untuk mengelompokan/memodelkan data set yang sudah tersedia, dan yang terakhir cluster distance performance yang digunakan untuk menguji kinerja hasil pengelompokan. Maka dalam penelitian ini, peneliti menetapkan 10 cluster/kelompok sesuai dengan nilai optimum yang telah dicoba.



Gambar 5 Data Mining

6. Interpretation / evaluation

Pada penelitian ini, cluster yang digunakan yaitu 10 cluster, dimulai dari cluster 0 sampai dengan cluster 9 dengan total dataset yang di clustering kan sebanyak 10 data.



Gambar 6 Interpretation/evaluation

Berdasarkan gambar 6 tentang cluster model menjelaskan bahwa dataset yang digunakan 10 record data. Terdiri dari:

- cluster 0 sebanyak 482 items,
- cluster 1 sebanyak 2 items,
- cluster 2 sebanyak 48 items,
- cluster 3 sebanyak 124 items,
- cluster 4 sebanyak 31 items,
- cluster 5 sebanyak 6 items,
- cluster 6 sebanyak 1 items,
- cluster 7 sebanyak 21 items,
- cluster 8 sebanyak 225 items,
- cluster 9 sebanyak 71 items.

B. Nilai DBI Hasil Pengelompokan Harga Rumah

Davies Bouldin atau nilai dbi ini bekerja untuk mengevaluasi kinerja dari proses pengklusteran, dimana dari evaluasi tersebut akan menghasilkan sebuah parameter berupa numeric. Ketika nilai yang diberikan semakin kecil, cluster yang diberikan semakin baik. Berikut ini proses mencari cluster optimum dimulai dari cluster 0 sampai dengan cluster 10.

Hasil penguraian proses mencari cluster optimum atau terbaik dari cluster 1 sampai cluster 10 dengan Davies Bouldin Index (DBI) menggunakan tools RapidMiner Studio sesuai dengan tabel 2.

Tabel 2 Nilai Dbi 1-10

Jumlah Cluster	Nilai DBI
0 Cluster	0.142
1 Cluster	0.142
2 Cluster	0.142
3 Cluster	0.166
4 Cluster	0.174
5 Cluster	0.167
6 Cluster	0.147
7 Cluster	0.167
8 Cluster	0.149
9 Cluster	0.130
10 Cluster	0.129

Hasil Dbi terbaik diperoleh dengan menggunakan cluster model 10 yang lebih kecil dan presisi, yang menghasilkan hasil 0,129. Jadi data processor akan menghasilkan cluster data terbaik apabila dikelompokkan menjadi 10 cluster.

Algoritma k-means clustering dapat mengelompokkan harga rumah di Jakarta Selatan dengan didapatkan pengklasteran terbaik dilakukan dengan 10 cluster, maka didapatkannya nilai dbi terbaik dari hasil pengelompokan harga rumah di Jakarta Selatan tersebut yaitu sebesar 0.129 berada pada cluster 10. Dari 10 record data maka diperoleh :

- 482 record data pada cluster pertama,
- 2 record data pada cluster kedua,
- 48 record data pada cluster ketiga,
- 123 record data pada cluster keempat,

- 31 record data pada cluster kelima,
- 6 record data pada cluster ke enam,
- 1 record data pada cluster ketujuh,
- 21 record data pada cluster kedelapan,
- 225 record data pada cluster kesembilan,
- 71 record data pada cluster kesepuluh.

KESIMPULAN DAN SARAN

Berdasarkan dari hasil penelitian dan pembahasan, maka diperoleh kesimpulan sebagai berikut:

1. Dari hasil clustering, maka diperoleh 10 kelompok cluster yaitu,
 - 482 record data pada cluster pertama,
 - 2 record data pada cluster kedua,
 - 48 record data pada cluster ketiga,
 - 123 record data pada cluster keempat,
 - 31 record data pada cluster kelima,
 - 6 record data pada cluster ke enam,
 - 1 record data pada cluster ketujuh,
 - 21 record data pada cluster kedelapan,
 - 225 record data pada cluster kesembilan,
 - 71 record data pada cluster kesepuluh.
2. Dari pencarian Nilai Dbi Hasil Mengelompokan Harga Rumah Menggunakan Algoritma K-Means Clustering dengan menggunakan sepuluh cluster maka menghasilkan nilai Davies Bouldin Index atau dbi sebesar 0.129,
 - rata rata dalam jarak pusat 0 : 267562051182893728.000,
 - rata rata dalam jarak pusat 1 : 45773.083,
 - rata rata dalam jarak pusat 2 : 791040262574103040.000,
 - rata rata dalam jarak pusat 3 : 271578686725526560.000,
 - rata rata dalam jarak pusat 4 : 497334388887154940.000,
 - rata rata dalam jarak pusat 5 : 1407407407407418110.000,
 - rata rata dalam jarak pusat 6 : - 0.000,
 - rata rata dalam jarak pusat 7 : 1728344671201840130.000,
 - rata rata dalam jarak pusat 8 : 217307962603634176.000,
 - rata rata dalam jarak pusat 9 : 470403085943535550.000.

Berdasarkan pengerjaan penelitian ini, tentunya masih banyak kekurangan. Maka dari itu peneliti memberikan saran untuk pengembangan lebih lanjut, yaitu:

1. Harus dilakukan proses text preprocessing yang jauh lebih baik agar tidak dapat data yang ambigu/duplikat, supaya data yang dihasilkan jauh lebih baik.
2. Harus dilakukan analisis yang lebih teliti terhadap penggunaan algoritma clustering dan sistem klasifikasi lain sebagai pembandingan untuk hasil yang lebih akurat atau lebih baik.

DAFTAR REFERENSI

- (Y. R. Sari et al., 2020) Al-rizki, M. F. I., Widaningrum, I., & Buntoro, G. A. (2020). *Prediksi Penyebaran Penyakit TBC dengan Metode K-Means Clustering Menggunakan Aplikasi Rapidminer*. 5(1), 1–10.
<https://doi.org/10.31544/jtera.v5.i1.2020.1-10>
- Apuilino Iman Seno Aji, F., Achmadi, S., & Ariwibisono, F. (2021). Penerapan Metode Clustering Pada Analisis Realisasi Pendapatan Asli Daerah Dengan Algoritma K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(2), 443–451.
<https://doi.org/10.36040/jati.v5i2.3741>
- Bastian, A., Sujadi, H., & Febrianto, G. (n.d.). *Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)*. 1, 26–32.
- Dinata, R. K., Safwandi, S., Hasdyna, N., & Azizah, N. (2020). Analisis K-Means Clustering pada Data Sepeda Motor. *INFORMAL: Informatics Journal*, 5(1), 10. <https://doi.org/10.19184/isj.v5i1.17071>
- Fatmawati, K., & Windarto, A. P. (2018). Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi. *Computer Engineering, Science and System Journal*, 3(2), 173. <https://doi.org/10.24114/cess.v3i2.9661>
- Ginting, F., Buulolo, E., & Siagian, E. R. (2019). Implementasi Algoritma Regresi Linear Sederhana Dalam Memprediksi Besaran Pendapatan Daerah (Studi Kasus: Dinas Pendapatan Kab. Deli Serdang). *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 274–279.
<https://doi.org/10.30865/komik.v3i1.1602>
- Mohede, R. M., Rotinsulu, D. C., Tumangkang, S. Y. L., Pembangunan, J. E., Ekonomi, F., & Ratulangi, U. S. (2020). *DAERAH TERHADAP PENINGKATAN PENDAPATAN ASLI DAERAH (PAD) DI KABUPATEN KEPULAUAN SANGIHE*. 20(01), 45–54.
- Purba, W., Siawin, W., & . H. (2019). Implementasi Data Mining Untuk Pengelompokkan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan

Algoritma K-Means Clustering. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, 2(2), 85–90.
<https://doi.org/10.34012/jusikom.v2i2.429>

Ramdhan, D., Dwilestari, G., Dana, R. D., Ajiz, A., & Kaslani, K. (2022). Clustering Data Persediaan Barang Dengan Menggunakan Metode K-Means. *MEANS (Media Informasi Analisa Dan Sistem)*, 7(1), 1–9.
<https://doi.org/10.54367/means.v7i1.1826>

Sagala, R. M. (2021). Prediksi Kelulusan Mahasiswa Menggunakan Data mining Prediction of college subject using K-means Algorithm in Data mining. *Jurnal TeIKa*, 11(2), 131–142.

Sari, R. M., Tasril, V., & M, Y. A. (2020). *Prediksi Jumlah APBD Kota Payakumbuh dengan metode K-Means*. 1, 45–50.

Sari, Y. R., Sudewa, A., Lestari, D. A., & Jaya, T. I. (2020). Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 192. <https://doi.org/10.24114/cess.v5i2.18519>

Studi, P., & Informatika, T. (2022). *Vol . 15 No . 1 September 2022 ISSN : 1979-8415 PENERAPAN ALGORITMA K-MEANS UNTUK MENGELOMPOKKAN DATA PENGIRIMAN PAKET DI KANTOR POS CIREBON ISSN : 1979-8415*. 15(1), 23–27.