

PENERAPAN ALGORITMA K-MEANS UNTUK CLUSTERING HARGA RUMAH DI BANDUNG

Briyan Gifari Aji¹⁾, Dwi Chandra Aditya Sondawa²⁾, Muhammad Rifky Gifari³⁾, Sena Wijayanto⁴⁾

^{1), 2), 3), 4)} Sistem Informasi, Institut Teknologi Telkom Purwokerto

Jl. DI Panjaitan No.128, Karangreja, Purwokerto Kidul, Banyumas, Jawa Tengah 53147

Email : 20103015@ittelkom-pwt.ac.id¹⁾, 20103018@ittelkom-pwt.ac.id²⁾, 20103161@ittelkom-pwt.ac.id³⁾, sena@ittelkom-pwt.ac.id⁴⁾

ABSTRACT

The need for shelter is one of the fundamental aspects of daily life for humans. A house serves not only as a place to seek protection and rest but also as a venue for socializing with family. One of the factors influencing the decision in choosing a house is its price. House prices vary in each region, depending on factors such as location and other attributes. In major cities like Bandung, house prices differ based on their categories. However, many people still find it challenging to determine the value and discern whether a house is classified as affordable or expensive. Hence, there is a need for a clustering process of house prices in Bandung to aid in comprehending and categorizing house prices based on attributes such as the house price, total building area, and total land area. To understand and analyze the patterns of house prices in Bandung, this study utilizes the K-Means method to cluster the house price data into several groups based on their similarity in attributes. Additionally, the research aims to determine the optimal number of clusters through the cluster validation process using the silhouette index. The findings show that when using $n_cluster=2$, a silhouette score of 0.8870 is obtained, and with $n_cluster=3$, the silhouette score is 0.8009. These results indicate that clustering with $n_cluster=2$ and $n_cluster=3$ both exhibit strong interpretative structures. Thus, the clustering of house prices in Bandung can be effectively grouped into 2 clusters, as evidenced by the higher silhouette score obtained with $n_cluster=2$, approaching 1 compared to $n_cluster=3$.

Keywords : house, price, clustering, K-Means, silhouette

ABSTRAK

Kebutuhan akan tempat tinggal merupakan salah satu aspek mendasar dalam kehidupan sehari-hari. Rumah bukan hanya berfungsi sebagai tempat untuk melindungi diri dan beristirahat, tetapi juga menjadi tempat bersosialisasi dengan keluarga. Dalam memilih sebuah rumah terdapat faktor yang mempengaruhi salah satunya adalah harga rumah. Harga rumah di setiap daerah bervariasi tergantung pada lokasi dan faktor lainnya. Di kota-kota besar seperti Bandung, harga rumah pun berbeda-beda sesuai dengan kategorinya. Namun, masih banyak masyarakat yang kesulitan dalam menentukan nilai dan mengetahui apakah harga rumah tersebut termasuk golongan murah atau mahal. Oleh karena itu, perlu adanya proses *clustering* harga rumah di Bandung untuk membantu pemahaman dan pengelompokan harga rumah berdasarkan atribut seperti harga rumah, total luas bangunan, dan total luas tanah. Untuk memahami dan menganalisis pola harga rumah di Bandung, dalam penelitian ini, metode K-Means digunakan untuk mengelompokkan data harga rumah menjadi beberapa *cluster* berdasarkan kesamaan atributnya. Selain itu, penelitian ini juga bertujuan untuk menentukan jumlah *cluster* yang optimal melalui proses cluster validation menggunakan indeks silhouette. Hasil penelitian menunjukkan bahwa saat menggunakan $n_cluster=2$, diperoleh *silhouette score* sebesar 0,8870 dan $n_cluster=3$ diperoleh *silhouette score* sebesar 0,8009. Hal ini menunjukkan bahwa pengelompokan dengan $n_cluster=2$ dan $n_cluster=3$, memiliki interpretasi struktur yang kuat. Oleh karena itu, pengelompokan harga rumah di Bandung dapat di kelompokkan menjadi 2 *cluster*, sesuai hasil *silhouette score* dari $n_cluster=2$ yang memberikan hasil lebih tinggi dan lebih mendekati 1 dari $n_cluster=3$.

Kata Kunci : rumah, harga, clustering, K-Means, silhouette



Article History

Received : 21/07/2023
Revised : 29/07/2023
Accepted : 30/07/2023
Online : 01/08/2023



This is an open access article under the
CC BY-SA 4.0 License

1. Pendahuluan

Dalam kehidupan sehari-hari, kebutuhan manusia yang mendasar mencakup berbagai aspek, termasuk tempat tinggal (Juniwati Ayuningtyas & Purwaning Astuti, 2018). Rumah merupakan salah satu kebutuhan pokok yang penting bagi manusia untuk melindungi diri, beristirahat, dan menjalani aktivitas sehari-hari. Selain berfungsi sebagai tempat tinggal yang mendasar, rumah juga memiliki peran penting sebagai tempat untuk bersosialisasi dengan keluarga dan sanak saudara (Afmi Afika & Ariusni, 2019). Suatu rumah menjadi tempat berkumpul dan bersilaturahmi, menciptakan hubungan yang erat dan kehangatan di antara anggota keluarga. Oleh karena itu, rumah menjadi kebutuhan yang sangat penting bagi setiap orang. Kenyamanan sebuah rumah menjadi salah satu faktor dalam memilih sebuah rumah. Aspek kenyamanan meliputi lingkungan sekitar, luas bangunan rumah, dan faktor internal seperti harmoni keluarga yang mampu menciptakan suasana yang nyaman di dalam rumah. Selain kenyamanan faktor lain yang menjadi pertimbangan utama dalam pemilihan rumah adalah harganya (Faradilla Daldiri - et al., 2022).

Penting bagi setiap orang untuk melakukan pertimbangan yang matang dalam memastikan apakah harga suatu rumah sesuai dengan kualitas dan fasilitas yang disediakan. Harga rumah dapat berbeda-beda secara signifikan tergantung pada berbagai faktor, seperti yang terlihat pada contohnya di kota Bandung, di mana harga rumah bervariasi berdasarkan tipe, luas tanah, dan luas bangunan rumah (Septiani et al., 2023). Faktor-faktor seperti luas tanah dan bangunan menjadi kriteria utama dalam menetapkan harga sebuah rumah. Faktor ini menjadi acuan dalam mengambil keputusan dalam membeli sebuah rumah sesuai dengan anggaran yang dimiliki. Oleh karena itu perlu dilakukan sebuah teknik untuk mengelompokkan harga rumah yang ada di Bandung untuk membedakan harga rumah pada setiap luas tanah dan bangunan rumah yang ada mulai dari harga terendah hingga harga tertinggi. Pengelompokan ini dapat dilakukan dengan menggunakan teknik *clustering*.

Clustering adalah metode pengelompokan data yang bertujuan untuk mengelompokkan objek yang serupa ke dalam kelompok-kelompoknya (Priyatman et al., 2019). Salah satu algoritma *clustering* yang sering digunakan adalah algoritma *K-Means*. Algoritma *K-Means* merupakan metode yang berbasis partisi, di mana data dikelompokkan ke dalam kelompok (*clusters*) berdasarkan jarak Euclidean antara titik data [cith1]. Algoritma *K-Means* merupakan algoritma yang hanya bekerja pada atribut numerik. Algoritma ini membutuhkan iterasi yang berulang untuk mencapai konvergensi di mana pusat kelompok dan anggota kelompok tidak berubah.

Pada penelitian sebelumnya, dalam mengelompokkan data obat-obatan di Rumah Sakit Umum Daerah Pekanbaru menggunakan algoritma *K-Means* (Gustientiedina et al., 2019). Hasil penelitian ini berupa pengelompokkan data obat berdasarkan *cluster*

yang telah dikategorikan secara efektif menggunakan algoritma *K-Means*, untuk memberikan referensi dalam perencanaan dan pengendalian pasokan medis di Rumah Sakit Umum Daerah Pekanbaru. Dengan demikian, penelitian ini dapat digunakan sebagai referensi metode untuk melakukan pengelompokkan terhadap daftar harga rumah di Bandung berdasarkan faktor-faktor numerik seperti luas bangunan, luas tanah, dan variabel lainnya guna mengetahui harga rumah dari tertinggi hingga terendah.

Tujuan dari penelitian ini adalah untuk menerapkan algoritma *K-Means* dalam melakukan proses *clustering* harga rumah di Bandung. Metode *K-Means* digunakan untuk mengelompokkan data harga rumah ke dalam beberapa *cluster* berdasarkan kesamaan atributnya. Selain itu, penelitian ini bertujuan untuk menentukan jumlah *cluster* yang optimal melalui proses *Cluster Validation*. Proses *Cluster Validation* dilakukan dengan menggunakan indeks *Silhouette* untuk mengukur kualitas dan kecocokan *clustering* yang dihasilkan oleh algoritma *K-Means*.

A. Rumah

Rumah merupakan suatu tempat yang menjadi wadah bagi manusia dan digunakan sebagai tempat tinggal beserta berbagai aktivitas yang dilakukan di dalamnya (Ade Wulandari et al., 2020). Rumah adalah sebuah bangunan fisik yang berfungsi sebagai tempat perlindungan, tempat tinggal, dan juga menjadi tempat beraktivitas bagi individu, kelompok, atau keluarga.

B. Clustering

Clustering merupakan teknik yang digunakan untuk mengelompokkan atau mengkategorikan sekelompok objek berdasarkan kesamaan atribut atau karakteristik dengan data lainnya (Hardiani, 2022). Metode ini termasuk dalam kategori *unsupervised*, yang berarti tidak memerlukan arahan atau target spesifik untuk menghasilkan kelompok-kelompok tersebut (Jayanta et al., 2020). Dalam *clustering*, tidak ada label atau penamaan khusus yang diberikan pada kelompok data yang terbentuk (Mauliadi, 2022).

C. Algoritma K-Means

Algoritma *K-Means* relatif sederhana dalam implementasinya mengandalkan data yang diperoleh selama proses dan penarikan kesimpulan yang akan dilakukan (Nurani et al., 2023). Metode ini membutuhkan pengaturan jumlah *cluster*, pemilihan pusat *cluster* awal (*centroid*), dan perhitungan jarak antara *centroid* dengan setiap objek data (Faradilla Daldiri - et al., 2022). Dalam algoritma *K-Means*, perlu dilakukan perhitungan jarak antara data dengan *centroid* terdekat menggunakan rumus:

$$d(i, j) = \sqrt{(X_{i1} + X_{j1})^2 + \dots + (X_{ip} + X_{jp})^2} \quad (1)$$

$d(i,j)$ = Jarak data ke centroid
 X_i = Record / Data
 X_i = Centroid / Pusat cluster

D. Metode Elbow

Metode *Elbow* adalah metode yang digunakan untuk menentukan jumlah *cluster* yang optimal dalam analisis cluster dengan algoritma *K-Means* (Dista & Abdulloh, 2022). Metode ini melibatkan menghitung *Within-Cluster Sum of Squares* (WCSS) untuk berbagai jumlah *cluster* dan menggambarkan perubahan WCSS dalam bentuk grafik. dengan mencari titik pada grafik yang menyerupai siku (Aditya et al., 2020). Untuk mendapatkan perbandingan dengan menghitung SSE (*Sum of Square Error*) dari masing-masing nilai *cluster*. Rumus SSE pada *K-Means* dapat dilihat pada persamaan 2 sebagai berikut:

$$SSE = \sum_{i=1}^n (x_i - \bar{x}) \tag{2}$$

E. Silhouette

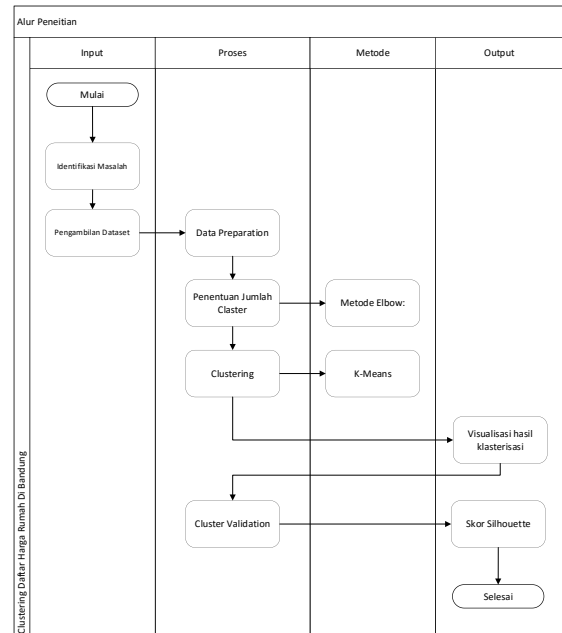
Dalam melakukan pengujian validitas *cluster* yang digunakan dapat diukur dengan validitas *Silhouette* yang melibatkan perhitungan rata-rata nilai pada setiap kelompok data. Perhitungan rata-rata ini didasarkan pada nilai kesamaan (*separation*) dan keseragaman (*compactness*) dari data dalam *cluster*. Nilai *Silhouette* mendekati 1 menunjukkan *cluster* yang baik dan valid. Tabel 1 memberikan rentang nilai dan interpretasi dengan ketentuan *Cluster* dengan nilai *Silhouette* mendekati 1 dianggap sebagai hasil *cluster* yang terbaik (Nurina Sari & Primajaya, 2019).

Tabel 1. Rentang Nilai Silhouette dan Interpretasi

Rentang Nilai	Interpretasi
0.71 – 1.0	Struktur kuat
0.51 – 0.70	Struktur yang beralasan
0.26 – 0.50	Struktur lemah
<0.25	Tidak ditemukan struktur yang substansial

2. Pembahasan

Dalam pembahasan ini menjelaskan tentang penggunaan algoritma *K-Means* pada *clustering* harga rumah di Bandung, visualisasi langkah-langkah proses *clustering K-Means* beserta *input*, proses, metode, dan *output*, dapat dilihat pada gambar 1.



Gambar 1. Alur Penelitian

Alur penelitian yang ada pada Gambar 1, dijelaskan sebagai berikut :

A. Identifikasi Masalah

Mengenali permasalahan yang sedang terjadi dan mencari sumber data yang relevan untuk dapat digunakan dalam penyelesaiannya.

B. Pengambilan Dataset

Pengumpulan *dataset* dilakukan dengan melakukan pengambilan data dari situs Kaggle <https://www.kaggle.com/datasets/rafliaping/dataset-harga-rumah-bandung>. Data yang di dapatkan akan dikelola menggunakan aplikasi *Jupyter Notebook*, Total data tersebut berjumlah 1470 data yang dapat dilihat pada Gambar 2.

index	judul	alamat	deskripsi	kamar	bangunan	lahan	harga
0	Promo Rumah Murah Akhir Tahun	Antapani, Bandung	Promo akhir tahun !! Hanya 500 j-an + FREE SH...	3	75	84	500000000
1	Rumah modern Readystock di Bandung rasa cilind...	Bandung Wetan, Bandung	Rumah Baru modern 2 LT di Saye Kota Cinahi	3	70	71	700000000
2	Dipai rumah 300 jutaan jalan lebar dekat lisa...	Padasirang, Bandung	!Jadilah konsumen CERDAS !! Tips singkat seke...	2	38	72	388000000
3	Dipai PROMO rumah 2 lantai di saye Setabuit...	Bandung	Copy selling nasec NUANSA ALAM SETABUHI CLOV...	2	60	60	499000000
4	Rumah Strategis Siap Huni di dekat Tol Cibiru...	Cibiru Hill, Bandung	Hunan Nyaman di Cibiru, Cibinuyi, Bandung Tim...	2	90	114	397500000
1465	Rumah Lama Turangga	Turangga, Bandung	Di jual rumah lama turangga! Luas tanah: 240 ...	6	280	240	280000000
1466	Rumah di nirwana residence bandung siap huni	Bandung	Rumah bagus siap huni dan terawat, dekat deng...	3	180	90	1000000000
1467	Rumah Minimalis Siap Huni Sangat Murah Di Pitu...	Soekarno Hatta, Bandung	Turun Hargeharga Awal 675 sekarang jadi 572 re...	2	65	72	572000000
1468	Dipai Cepat Rumah Strategis di Komplek Sibar...	Rancasan, Bandung	Dipai Cepat Rumah di Komplek Batu Raden deka...	6	160	318	2300000000
1469	Rumah turangga bagus siap huni harga dila...	Turangga, Bandung	Turun Harga dr awal CAKEP BOSQUE Area Turan...	5	240	188	2400000000

Gambar 2. Data Harga Rumah di Bandung

Berdasarkan Gambar 2, data tersebut memiliki 7 atribut diantaranya; judul : judul penjualan rumah, alamat : lokasi atau letak rumah berada, deskripsi : penjelasan singkat tentang rumah, kamar : jumlah kamar yang tersedia, bangunan : total luas bangunan, lahan : total luas tanah, dan harga : harga dari rumah. Informasi detail terkait *dataset* tersebut dapat ditemukan pada Gambar 3.

```
In [3]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   judul       1470 non-null   object
1   alamat      1470 non-null   object
2   deskripsi   1468 non-null   object
3   kamar       1470 non-null   int64
4   bangunan    1470 non-null   int64
5   lahan       1470 non-null   int64
6   harga       1470 non-null   int64
dtypes: int64(4), object(3)
memory usage: 80.5+ KB
```

Gambar 3. Informasi Detail Dataset

Sedangkan statistik deskriptif terkait dataset dapat dilihat pada Gambar 4.

```
In [4]: df.describe()

Out[4]:
```

	kamar	bangunan	lahan	harga
count	1470.000000	1470.000000	1470.000000	1.470000e+03
mean	3.646259	177.489116	210.235374	2.581586e+09
std	2.535528	176.726326	500.109095	4.358051e+09
min	0.000000	0.000000	0.000000	6.500000e+07
25%	2.000000	65.000000	80.000000	6.750000e+08
50%	3.000000	122.500000	120.000000	1.370000e+09
75%	4.000000	230.000000	210.000000	3.000000e+09
max	47.000000	1850.000000	15000.000000	5.700000e+10

Gambar 4. Statistik Deskriptif

C. Data Preparation

Data preparation digunakan untuk melakukan persiapan dan transformasi data sehingga menjadi lebih sesuai dan siap digunakan dalam proses analisis atau pemodelan.

1. Data Cleaning

Melakukan pembersihan data untuk mengidentifikasi data yang memiliki nilai kosong atau bernilai null, serta mendeteksi adanya duplikasi dalam dataset. Selama proses ini, perlu dilihat komposisi persentase dari nilai yang hilang (missing value) yang terdapat pada dataset yang dapat dilihat pada Gambar 5.

```
In [4]: df_null = round(100*(df.isnull().sum())/len(df), 2)
df_null

Out[4]: judul      0.00
alamat    0.00
deskripsi 0.14
kamar     0.00
bangunan  0.00
lahan     0.00
harga     0.00
dtype: float64
```

Gambar 5. Data Harga Rumah di Bandung

Berdasarkan Gambar 5, dapat disimpulkan bahwa hanya satu atribut dalam dataset yang memiliki missing value. Langkah selanjutnya adalah melakukan penghapusan terhadap baris-baris yang mengandung missing value. Proses ini bertujuan untuk membersihkan data dari baris yang tidak lengkap sehingga dataset

menjadi lebih bersih dan siap untuk digunakan dalam analisis lebih lanjut, seperti yang terlihat pada Gambar 6.

```
In [5]: df = df.dropna()
df.shape

Out[5]: (1468, 7)
```

Gambar 6. Penghapusan Data

Pada Gambar 6, jumlah dataset yang awalnya 1470 berkurang menjadi 1468 setelah di bersihkan data yang mengandung missing value

2. Data Transformation

Melakukan transformasi pada data, dengan membuat sebuah kolom baru bernama 'luas total'. Kolom 'luas total' diperoleh dengan mengalikan kolom 'lahan' dan 'bangunan', yang dapat dilihat pada Gambar 7.

```
In [6]: df['luas_total'] = df['lahan'] * df['bangunan']
df[['harga', 'luas_total']]

Out[6]:
```

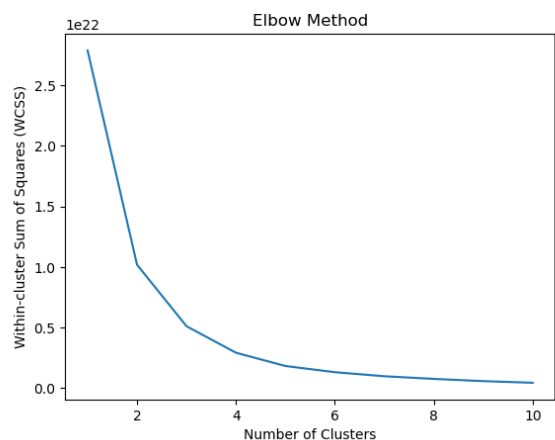
	harga	luas_total
0	590000000	6300
1	700000000	4970
2	386000000	2592
3	499000000	3600
4	397500000	10260
...
1465	2800000000	67200
1466	1000000000	16200
1467	572000000	4680
1468	2300000000	50880
1469	2400000000	45120

1468 rows x 2 columns

Gambar 7. Pembuatan Kolom Baru

D. Penentuan Jumlah Cluster

Melakukan penentuan jumlah cluster yang optimal menggunakan metode Elbow. Penentuan jumlah cluster ini yang dapat dilihat pada Gambar 8.



Gambar 8. Penentuan Jumlah Cluster

Pada Gambar 8, diketahui jika terdapat 2 *number cluster*, yang menyerupai siku yaitu pada $n_cluster=2$ dan $n_cluster=3$.

E. Clustering

Number cluster yang sudah diperoleh, kemudian dikelompokkan berdasarkan *cluster* (*k*) masing-masing dengan menggunakan algoritma *K-Means*.

1. $n_cluster=2$

Selanjutnya, dilakukan perhitungan untuk melihat nilai jarak objek data ke centroid pada $n_cluster=2$, dengan menggunakan array, serupa dengan yang ditampilkan pada Gambar 9.

```
In [9]: kmeans = KMeans(n_clusters=2, init='k-means++', max_iter=50)
kmeans.fit(X)
distances = kmeans.transform(X)
print(distances)

[[1.41128987e+09 2.27475000e+10]
 [1.30128987e+09 2.26375000e+10]
 [1.61528987e+09 2.29515000e+10]
 ...
 [1.42928987e+09 2.27655000e+10]
 [2.98710133e+08 2.10375000e+10]
 [3.98710133e+08 2.09375000e+10]]
```

Gambar 9. Jarak Objek Data ke Centroid $n_cluster=2$

Mengelompokkan data $n_cluster=2$, dengan menggunakan algoritma *K-Means*, dapat dilihat pada Gambar 10.

```
In [10]: y_kmeans2 = kmeans.fit_predict(X)
y_kmeans2

Out[10]: array([0, 0, 0, ..., 0, 0, 0])
```

Gambar 10. *K-Means Label* pada $n_cluster=2$

Hasil dari algoritma *K-Means* pada data $n_cluster=2$ yang telah memisahkan titik-titik data menjadi kelompok-kelompok yang berbeda, dapat dilihat pada Gambar 11.

```
In [11]: X['class'] = y_kmeans2
X

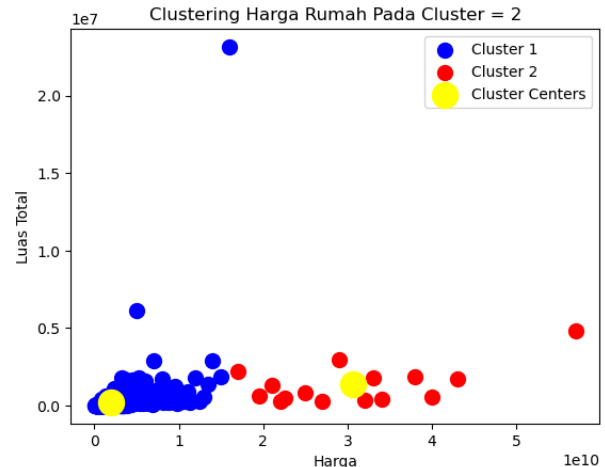
Out[11]:
```

	harga	luas_total	class
0	590000000	6300	0
1	700000000	4970	0
2	386000000	2592	0
3	499000000	3600	0
4	397500000	10260	0
...
1465	2800000000	67200	0
1466	1000000000	16200	0
1467	572000000	4680	0
1468	2300000000	50880	0
1469	2400000000	45120	0

1468 rows x 3 columns

Gambar 11. Data Hasil *K-Means* pada $n_cluster=2$

Pengelompokan data berdasarkan *class* pada $n_cluster=2$ yang telah diperoleh, akan diaplikasikan ke dalam bentuk *scatter plot* seperti pada Gambar 12.



Gambar 12. Pengelompokan Data pada $n_cluster=2$

Pada Gambar 12, data dikelompokkan menjadi 2 *cluster*, dimulai dari terendah sampai tertinggi secara berurutan yaitu *blue* dan *red*. Untuk melihat jumlah data pada setiap *class* yang ada pada $n_cluster=2$, dapat dilihat pada Gambar 13.

```
In [12]: X['class'].value_counts()

Out[12]: 0    1426
         1     42
         Name: class, dtype: int64
```

Gambar 13. Total Data Setiap *Class* pada $n_cluster=2$

Pada Gambar 13, terlihat informasi tentang jumlah titik data pada masing-masing *class* yang terdapat pada $n_cluster=2$. Terdapat total 2 *class*, yang diurutkan dari yang terbanyak hingga terkecil. Jumlah data pada masing-masing *class* yaitu *class 0* memiliki 1426 data dan *class 1* memiliki 42 data.

2. $n_cluster=3$

Melakukan perhitungan untuk melihat nilai jarak objek data ke *centroid* pada $n_cluster=3$, dengan menggunakan array, serupa dengan yang ditampilkan pada Gambar 14.

```
In [13]: kmeans = KMeans(n_clusters=3, init='k-means++', max_iter=50)
kmeans.fit(X)
distances = kmeans.transform(X)
print(distances)

[[1.08707725e+09 3.61023077e+10 1.03870192e+10]
 [9.77077245e+08 3.59923077e+10 1.02770192e+10]
 [1.29107725e+09 3.63063077e+10 1.05910192e+10]
 ...
 [1.10507725e+09 3.61203077e+10 1.04050192e+10]
 [6.22922755e+08 3.43923077e+10 8.67701916e+09]
 [7.22922755e+08 3.42923077e+10 8.57701916e+09]]
```

Gambar 14. Jarak Objek Data ke Centroid $n_cluster=3$

Mengelompokkan data $n_cluster=3$, dengan menggunakan algoritma *K-Means*, dapat dilihat pada Gambar 15.

```
In [15]: y_kmeans3 = kmeans.fit_predict(X)
y_kmeans3
Out[15]: array([0, 0, 0, ..., 0, 0, 0])
```

Gambar 15. K-Means Label pada n_cluster=3

Hasil dari algoritma K-Means pada data n_cluster=3 yang telah memisahkan titik-titik data menjadi kelompok-kelompok yang berbeda, dapat dilihat pada Gambar 16.

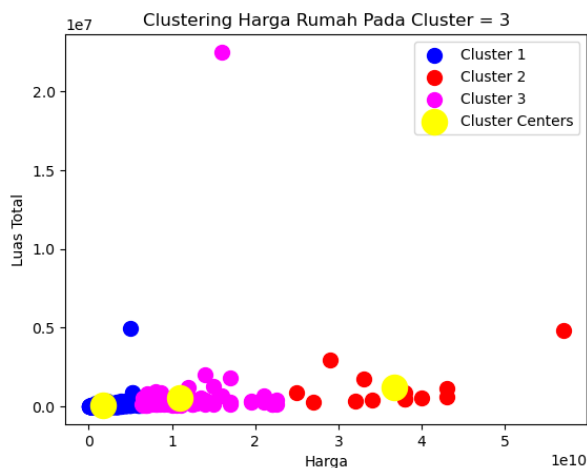
```
In [16]: X['class'] = y_kmeans3
X
Out[16]:
```

	harga	luas_total	class
0	5900000000	6300	0
1	7000000000	4970	0
2	3860000000	2592	0
3	4990000000	3600	0
4	3975000000	10260	0
...
1465	28000000000	67200	0
1466	10000000000	16200	0
1467	5720000000	4680	0
1468	23000000000	50880	0
1469	24000000000	45120	0

1468 rows x 3 columns

Gambar 16. Data Hasil K-Means pada n_cluster=3

Pengelompokan data berdasarkan class pada n_cluster=3 yang telah diperoleh, akan diaplikasikan ke dalam bentuk scatter plot seperti pada Gambar 17.



Gambar 17. Pengelompokan Data pada n_cluster=3

Pada Gambar 17, data dikelompokkan menjadi 3 cluster, dimulai dari terendah sampai tertinggi secara berurutan yaitu blue, purple dan red. Untuk melihat jumlah data pada setiap class yang ada pada n_cluster=3, dapat dilihat pada Gambar 18.

```
In [17]: X['class'].value_counts()
Out[17]: 0    1361
         2     94
         1     13
Name: class, dtype: int64
```

Gambar 18. Total Data Setiap Class pada n_cluster=3

Pada Gambar 18, terlihat informasi tentang jumlah titik data pada masing-masing class yang terdapat pada n_cluster=3. Terdapat total 3 class, yang diurutkan dari yang terbanyak hingga terkecil. Jumlah data pada masing-masing class yaitu class 0 memiliki 1361 data, class 2 memiliki 94 data dan class 1 memiliki 13 data.

F. Cluster Validation

Cluster validation digunakan untuk mengevaluasi kualitas dan kecocokan clustering yang dihasilkan oleh algoritma K-Means. Mengevaluasi seberapa baik kualitas clustering yang dihasilkan oleh algoritma K-Means menggunakan Silhouette Scoor. Silhouette Scoor pada n_cluster=2, dapat dilihat pada Gambar 19.

```
In [24]: sscore_kmeans = silhouette_score(X, y_kmeans2)
sscore_kmeans
Out[24]: 0.8870355385990925
```

Gambar 19. Silhouette Score pada n_cluster=2

Pada Gambar 19, diperoleh Silhouette Score pada n_cluster=2, sebesar 0,8870, yang berarti hasil tersebut memiliki interpretasi hasil yang memiliki struktur yang tinggi. Sedangkan Silhouette Score pada n_cluster=3, dapat dilihat pada Gambar 17.

```
In [43]: sscore_kmeans = silhouette_score(X, y_kmeans3)
sscore_kmeans
Out[43]: 0.8009614510211441
```

Gambar 17. Silhouette Score pada n_cluster=3

Pada Gambar 17, diperoleh Silhouette Score pada n_cluster=3, sebesar 0,8009, yang berarti hasil tersebut memiliki interpretasi hasil yang memiliki struktur yang tinggi. Tabel perbandingan hasil Silhouette Score antar setiap cluster dapat dilihat pada Tabel 2.

Tabel 2. Perbandingan Silhouette Score

Cluster	Score	Interpretasi
2	0,8870	Struktur Kuat
3	0,8009	Struktur Kuat

3. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa dalam *clustering* daftar harga rumah di Bandung menggunakan algoritma *K-Means*, saat menggunakan $n_cluster=2$, diperoleh *silhouette score* yang lebih tinggi dan mendekati 1, yaitu sebesar 0,8870. Hal ini menunjukkan bahwa *clustering* dengan 2 *cluster* memiliki interpretasi struktur yang lebih kuat, karena data-data dalam cluster tersebut sangat berdekatan dan terpisah dengan baik dari *cluster* lainnya. Di sisi lain, ketika menggunakan $n_cluster=3$, diperoleh *silhouette score* sebesar 0,8009. Meskipun nilai *silhouette score* pada $n_cluster=3$, juga menunjukkan struktur yang kuat, namun nilai ini lebih rendah dibandingkan dengan *clustering* dengan $n_cluster=2$. Sehingga, dapat disimpulkan bahwa *clustering* harga rumah di Bandung dapat dikelompokkan menjadi 2 *cluster* sesuai dengan *silhouette score* sebesar 0,8870, yang menunjukkan hasil yang lebih optimal dan memberikan interpretasi struktur data yang lebih baik

Daftar Pustaka

- Ade Wulandari, S., Karmilah, M., & Yuliani, E. (2020). Perubahan Fungsi Rumah Menjadi Ruang Ekonomi pada Penggal Jalan Kawasan Wisata Pantai Kartini. *Sultan Agung Fundamental Research Journal*, 1, 44–53.
- Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(1), 51–58. <https://doi.org/10.30865/mib.v4i1.1784>
- Afmi Afika, Y., & Ariusni. (2019). Faktor-Faktor Yang Mempengaruhi Permintaan Rumah Di Indonesia. *Jurnal Kajian Ekonomi Dan Pembangunan*, 1, 497–508.
- Dista, T. M., & Abdulloh, F. F. (2022). Clustering Pengunjung Mall Menggunakan Metode K-Means dan Particle Swarm Optimization. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(3), 1339. <https://doi.org/10.30865/mib.v6i3.4172>
- Faradilla Daldiri -, Z., Rafly -, M., & Veritawati -, I. (2022). Clustering Daftar Harga Rumah di Jakarta Dengan Algoritma K-Means. *Journal of Informatics and Advanced Computing (JIAC)*, 3(2). <https://www.kaggle.com/datasets/wisnuanggara/daftar-harga-rumah-di-jakarta>
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>
- Hardiani, T. (2022). Analisis Clustering Kasus Covid 19 di Indonesia Menggunakan Algoritma K-Means. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 11(2), 156–165. <https://doi.org/10.23887/janapati.v11i2.45376>
- Jayanta, Sekarani, F. H., Khansa, S., & Ashari, S. (2020). Klustering Jumlah Penduduk Kota Bandung Per Kelurahan Dan Kecamatan Berdasarkan Status Perkawinan Pada Tahun 2019 Dengan Metode K-Means. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*.
- Juniwati Ayuningtyas, F., & Purwaning Astuti, I. (2018). Faktor Penentu Permintaan Rumah Tinggal Di Provinsi Daerah Istimewa Yogyakarta. *Jurnal Ekonomi & Studi Pembangunan*, 19(1). <https://doi.org/10.18196/jesp.19.1.3890>
- Mauliadi, R. (2022). Data Mining Menggunakan Algoritma K-Means Clustering dalam Analisis Tingkat Potongan Harga Terhadap Harga Jual Sepeda Motor Honda. *Jurnal Informatika Ekonomi Bisnis*, 4, 124–129. <https://doi.org/10.37034/infec.v4i4.156>
- Nurani, S., Syahra, Y., & Calam, A. (2023). Penerapan Data Mining Dalam Clustering Pencapaian Target Penjualan Menggunakan Algoritma K-Means. 2, 355–363. <https://ojs.trigunadharma.ac.id/index.php/jsi>
- Nurina Sari, B., & Primajaya, A. (2019). Penerapan Clustering Dbscan Untuk Pertanian Padi Di Kabupaten Karawang. *Jurnal Informatika Dan Komputer*, 4(1), 28–34. www.mapcoordinates.net/en
- Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 5.
- Septiani, N., Anwar, S., & Herdiana, R. (2023). Penerapan Algoritma K-Means Clustering Untuk Harga Rumah di Jakarta Selatan. *Trending: Jurnal Ekonomi, Akuntansi Dan Manajemen*, 1(2).