



Implementation of K-Means and Agglomerative Hierarchical Methods to House Clusterization

Abdul Rozaq

Fakultas Teknik, Program Studi Teknik Informatika, Universitas PGRI Madiun, Madiun, Indonesia

Email: rozaq@unipma.ac.id

Email Penulis Korespondensi: rozaq@unipma.ac.id

Abstrak—Masyarakat secara umum akan berpendapat jika membangun rumah dengan luas bangunan lebih besar akan menghabiskan biaya lebih mahal dari pada membangun rumah dengan luas bangunan lebih kecil. Pandangan tersebut tidak selalu benar karena bebaapa faktor yang mempengaruhi besar kecilnya biaya tergantung dari luas bangunan, perbandingan campuran antara semen, pasir dan kapur, kualitas bahan bangunan yang digunakan sehingga akan mempengaruhi biaya yang harus dipersiapkan. Berdasarkan kasus tersebut maka peneliti mengelompokkan data (*clustering*) menggunakan *euclidian distance* untuk mengukur jarak antar titik. Mengelompokkan 200 data berdasarkan 2 fitur yaitu fitur jumlah semen dan jumlah biaya yang belum punya label. Hasil penelitian menunjukkan *clustering* dengan metode *K-Means* mampu mengelompokkan 200 data kedalam 3 kelompok dengan hasil kelompok satu sebanyak 50 data, kelompok 2 sebanyak 50 data, kelompok 3 sebanyak 100 data dengan waktu komputasi sebesar 0,444 detik dan *silhouette* 0,82 sedangkan hasil penelitian *clustering* menggunakan metode *Agglomerative Hierarchical* dengan *single linkage* menunjukkan data dikelompok 1 sebanyak 100 data, kelompok 2 sebanyak 50 data, kelompok 3 sebanyak 50 data dengan waktu komputasi sebesar 3,22 detik dan *silhouette* 0,51

Kata Kunci: K-Means; Hierarchical; Agglomerative; Single Linkage; Silhouette

Abstract—People in general will think that building a house with a larger building area will cost more than building a house with a smaller building area. This view is not always correct because several factors that affect the size of the cost depend on the size of the building, the ratio of the mixture between cement, sand and lime, the quality of the building materials used so that it will affect the costs that must be prepared. Based on this case, the researchers grouped the data (*clustering*) using Euclidean distance to measure the distance between points. Grouping 200 data based on 2 features, namely the amount of cement and the amount of costs that do not have a label. The results showed that clustering with the K-Means method was able to group 200 data into 3 groups with the results of group one as much as 50 data, group 2 as much as 50 data, group 3 as much as 100 data with a computation time of 0.444 seconds and silhouette 0.82 while the results of clustering research using the Agglomerative Hierarchical method with a single linkage shows 100 data in group 1, 50 data in group 2, 50 data in group 3 with a computation time of 3.22 seconds and silhouette 0.51

Keywords: K-Means; Hierarchical; Agglomerative; Single Linkage; Silhouette

1. PENDAHULUAN

Pembangunan rumah hunian sangat diperlukan kalkulasi yang tepat sehingga material ataupun biaya dapat dipersiapkan oleh pemilik rumah atau jasa pengembang perumahan [1]. Banyak faktor yang mempengaruhi biaya rumah [2] diantaranya kontur tanah, luas bangun, material dari kebutuhan semen hingga cat untuk finishing dan biaya tenaga kerja. Pada penelitian ini peneliti akan mengelompokkan jenis rumah berdasarkan jumlah semen yang dibutuhkan untuk membuat rumah dan total biaya yang dikeluarkan dikarenakan kebutuhan semen dan total biaya sangat bervariasi tergantung dengan luas bangunan dan komposisi campuran material.

Luas bangunan lebih kecil belum tentu biayanya akan lebih murah daripada luas bangunan yang lebih besar, misal luas bangunan rumah 36meter persegi belum tentu akan lebih murah jika dibandingkan dengan luas bangunan 42meter persegi. Salah satu faktor yaitu perbandingan campuran antara semen, pasir dan kapur sehingga dengan variasi campuran tersebut akan sangat berpengaruh terhadap biaya yang dibutuhkan. Semakin sedikit semen dapat pembangunan rumah akan semakin sedikit biaya yang akan dikeluarkan tetapi semakin kecil semen dalam campuran maka kekuatan rumah akan semakin berkurang [3]. Bagaimana pengelompokkan bahan bangunan berdasarkan semen dan biaya dapat mengoptimalkan jumlah dana masyarakat untuk membangun rumah?

Berdasarkan banyak variasi dalam pencampuran komposisi adukan maka peneliti mengelompokkan data kebutuhan rumah berdasarkan banyaknya semen serta banyaknya biaya yang dikeluarkan menggunakan metode *K-Means* dan *Hierarchical*. Alasan peneliti menggunakan 2 fitur adalah komposisi adukan mewakili banyaknya variasi dalam adukan sehingga campuran semen akan menentukan biaya sedangkan biaya mewakili bahan bangunan yang jika ukuran bangunan lebih besar pasti membutuhkan bahan lebih banyak seperti kayu, reng, cat dan lainnya.

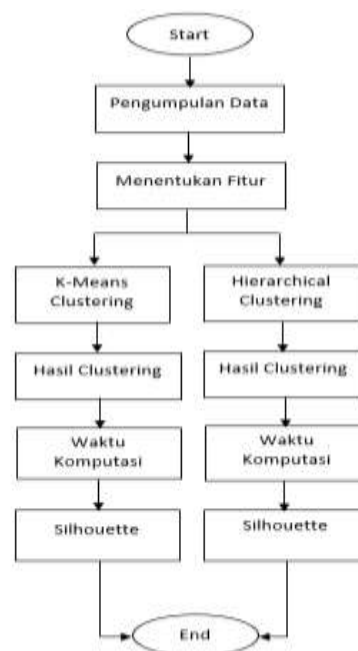
Beberapa rujukan jurnal terdahulu yang relevan dengan penelitian ini diantaranya Dalam penelitian [4] dari data 10 kecamatan dengan membagi menjadi 3 cluster menggunakan metode *K-Means* didapatkan hasil untuk klaster satu terdapat lima kecamatan, klaster dua sebanyak dua kecamatan dan klaster tiga sebanyak tiga kecamatan. Dalam penelitian [5] menghasilkan obat yang masuk kelompok *low* pemakaian pertahunnya di bawah 18000 biji, pemakaian obat *medium* pertahunnya diatas 18000 biji sedang pemakaian obat *high* pertahunnya mencapai diatas 70000 biji. Dalam penelitian [6] menyatakan pengelompokkan berdasarkan penjualan menjadi 3 cluster yaitu *low*, *medium* dan *high* kemudian data manual clustering dibandingkan dengan aplikasi menggunakan *K-Means* hasil yang didapat sebesar 100%. Dalam penelitian [7] menyatakan dengan dataset sebesar 10.773 dikelompokkan



menggunakan *K-Means* dengan 3 kelompok sedikit, sedang dan banyak peminat. Hasilnya kelompok sedikit sebesar 167 unit, kelompok sedang sebesar 2.803 unit dan kelompok banyak peminat sebesar 7.794 unit. Dalam penelitian [8] menyatakan dengan *Hierarchical clustering* untuk mengelompokkan data kredit di bank dengan pengujian sebesar 822 data maka diperoleh kelompok 1 sebesar 98%, kelompok 2 sebesar 1% dan kelompok 3 sebesar 1%. Dalam penelitian [9] menyatakan penjualan LINKAJA menggunakan *K-Means* dan *Hierarchical* memperoleh tingkat akurasi sebesar 93.33% untuk penjualan ditahun 2017 dan 90% untuk data penjualan ditahun 2018. Dalam penelitian [10] mengelompokkan pelanggan dengan 7 kelompok menggunakan *AHC* dengan hasil kelompok pertama sebanyak 20 data, kelompok dua sebanyak 43 data, kelompok tiga sebanyak 75 data, kelompok empat sebanyak 158 data, kelompok lima sebanyak 9 data, kelompok enam sebanyak 2 data dan kelompok tujuh sebanyak 1 data. Dalam penelitian [11] dengan sepuluh objek kabupaten dan sepuluh fitur menggunakan *AHC* dengan metode pengukuran *euclidean distance* dan *Manhattan distance* serta metodologi CRISP-DM dapat menghasilkan tingkat kemisikinan yang mudah dipahami dan sistematis. Dalam penelitian [12] dengan menggunakan tiga metode yang ada dalam *AHC* dapat menghasilkan dendrogram paling seimbang adalah *Average Linkage*

2. METODOLOGI PENELITIAN

Langkah-langkah dalam penelitian ditunjukkan pada gambar 1



Gambar 1. Langkah Penelitian

1. Pengumpulan data: pada tahap ini penelitian mengumpulkan data dengan cara observasi ke lapangan sehingga peneliti bisa memperoleh data sebanyak 200 data.
2. Menetuan fitur: data yang telah diperoleh terdapat 15 fitur sehingga peneliti harus memilih fitur yang paling cocok untuk di clustering sehingga terbentuk point yang bisa mewakili data maka penulis mengelompokkan berdasarkan jumlah semen dan biaya yang dibutuhkan.
3. Menerapkan data yang sudah diperoleh ke dalam metode *K-Means* dan *Hierarchical*.
4. Menghitung hasil clustering dengan *K-means* dan *Hierarchical*
5. Menghitung waktu yang diperlukan dari metode *K-Means* dan *Hierarchical*
6. Menghitung silhouette dari metode *K-Means* dan *Hierarchical*

2.1 Clustering

Clustering merupakan teknik membagi populasi atau titik data menjadi beberapa kelompok sehingga titik data pada kelompok yang memiliki kemiripan dengan titik data lain dalam kelompok yang sama dan berbeda dengan titik data pada kelompok lain. *Clustering* termasuk dalam pembelajaran tidak terawasi atau *unsupervised learning* [13].

2.2 K-Means

K-Means merupakan algoritma yang mengelompokkan yang sangat familiar dan mudah untuk diimplementasikan. *K-Means* membagi sekumpulan sampel kedalam k kelompok atau cluster yang terpisah menggunakan nilai rata-rata anggota sebagai indikator utama. Titik ini biasanya disebut Centroid, mengacu pada entitas aritmatika dengan nama yang sama, dan direpresentasikan sebagai vektor dalam ruang dimensi yang berubah-ubah. Setelah menetapkan titik



centroid maka alangkah selanjutnya menghitung jarak terdekat antar titik dengan menggunakan euclidean distance dengan persamaan (1)

$$d_{ij} = \sqrt{(a_i - a_j)^2 + (b_i - b_j)^2} \quad (1)$$

Dalam *K-Means* untuk menentukan banyaknya cluster yang dapat mempresentasikan data dengan baik maka dapat menggunakan metode Elbow [14]. Langkah-langkah *K-Means* ditunjukkan sebagai berikut:

1. sampel data yang tidak diklasifikasikan dan mengambil elemen k sebagai sentroid awal.
2. Menghitung jarak antara sampel data dan sampel yang dipilih pertama dan mendapatkan sentroid pertama yang dihitung.
3. Setelah sentroid berubah, perpindahannya akan memicu perubahan jarak individu, sehingga keanggotaan cluster dapat berubah.
4. Jika langkah 3 ada perubahan sentroid maka ulangi langkah 2.
5. Jika langkah ke 3 tidak ada perubahan sentroid maka proses selesai

2.3 Agglomerative Hierarchical

Pada prinsipnya hierarchical sama dengan pengelompokkan dengan *K-Means* yaitu pengelompokkan tanpa memberikan label pada data tetapi ada perbedaan yaitu setiap data adalah kluster sehingga data tersebut akan di bagi-bagi kedalam kluster sampai membentuk satu kluster kemudian dipecah kedalam jumlah kluster yang telah ditentukan. Langkah-langkah Hierarchical sebagai berikut:

1. Setiap point atau data merupakan kluster
2. Tentukan 2 poin yang saling berdekatan sehingga poin tersebut akan menjadi 1 kluster, misal diberi nama kluster A
3. Tentukan nilai yang paling dekat dengan kluster A.
4. Iterasi langkah kedua sampai hanya menghasilkan 1 kluster.

2.4 Single Linkage

Merupakan salah satu metode yang ada di *agglomerative* dengan mencari jarak terdekat dari kedua obyek dengan persamaan 1 kemudian hasil perhitungan dari persamaan 1 dipilih nilai terkecil dari jarak kedua obyek tersebut [15]

2.5 Silhouette Coefficient

Silhouette Coefficient menurut [16] menghitung nilai rerata setiap titik pada keseluruhan obyek yang ada dalam cluster dengan menggunakan persamaan (2)

$$S_i = \frac{b_i - a_i}{\max a_i, b_i} \quad (2)$$

Clustering dengan *K-Means*

3. HASIL DAN PEMBAHASAN

3.1 Penerapan K-Means

Dataset awal ditunjukkan pada tabel 1 merupakan data yang didapat dari pengembang perumahan.

Tabel 1. Dataset Awal

No	Semen (sak)	Biaya (Rp)
1	223,57	73.119.168
2	223,57	71.749.848
3	223,57	75.085.368
4	223,57	69.543.768
5	223,57	69.097.368
...
90	165,86	91.353.782,40
91	154,33	98.253.072
92	154,33	96.587.472
93	154,33	100.615.872
94	154,33	93.985.272
...
196	455,82	170.337.264,40
197	455,82	167.780.184,40
188	455,82	175.758.264,40



199	455,82	167.237.664,40
200	455,82	166.326.264,40

Tentukan cluster secara acak seperti pada tabel 2.

Tabel 2. Menentukan cluster awal

C1	117,44	65.786.287
C2	220,05	121.989.400
C3	799,93	185.500.072

Iterasi 1

Perhitungan menggunakan persamaan (1) dan data pada tabel 1.

$$\begin{aligned}
 d_{1,1} &= \sqrt{\frac{223,57^2 - 117,44^2 + 73.119.168^2}{65.786.287^2}} = 7.332.880,80 \\
 d_{1,2} &= \sqrt{\frac{223,57^2 - 220,05^2 + 73.119.168^2}{121.989.400^2}} = 48.870.232 \\
 d_{1,3} &= \sqrt{\frac{223,57^2 - 799,93^2 + 73.119.168^2}{185.500.072^2}} = 112.380.904 \\
 d_{2,1} &= \sqrt{\frac{223,57^2 - 117,44^2 + 71.749.848^2}{65.786.287^2}} = 5.963.560,80 \\
 d_{2,2} &= \sqrt{\frac{223,57^2 - 220,05^2 + 71.749.848^2}{121.989.400^2}} = 50.239.552 \\
 d_{2,3} &= \sqrt{\frac{223,57^2 - 799,93^2 + 71.749.848^2}{185.500.072^2}} = 113.750.224 \\
 d_{3,1} &= \sqrt{\frac{223,57^2 - 117,44^2 + 75.085.368^2}{-65.786.287^2}} = 9.299.080,80 \\
 d_{3,2} &= \sqrt{\frac{223,57^2 - 220,05^2 + 75.085.368^2}{-121.989.400^2}} = 46.904.032 \\
 d_{3,3} &= \sqrt{\frac{223,57^2 - 799,93^2 + 75.085.368^2}{-185.500.072^2}} = 110.414.704 \\
 d_{4,1} &= \sqrt{\frac{223,57^2 - 117,44^2 + 69.543.768^2}{-65.786.287^2}} = 3.757.480,80 \\
 d_{4,2} &= \sqrt{\frac{223,57^2 - 220,05^2 + 69.543.768^2}{-121.989.400^2}} = 52.445.632 \\
 d_{4,3} &= \sqrt{\frac{223,57^2 - 799,93^2 + 69.543.768^2}{-185.500.072^2}} = 115.956.304 \\
 d_{5,1} &= \sqrt{\frac{223,57^2 - 117,44^2 + 69.097.368^2}{-65.786.287^2}} = 3.311.080,80 \\
 d_{5,2} &= \sqrt{\frac{223,57^2 - 220,05^2 + 69.097.368^2}{-121.989.400^2}} = 52.892.032 \\
 d_{5,3} &= \sqrt{\frac{223,57^2 - 799,93^2 + 69.097.368^2}{-185.500.072^2}} = 116.402.704 \\
 d_{90,1} &= \sqrt{\frac{165,86^2 - 117,44^2 + 91.353.782,40^2}{-65.786.287^2}} = 26.106.895,20 \\
 d_{90,2} &= \sqrt{\frac{165,86^2 - 220,05^2 + 91.353.782,40^2}{-121.989.400^2}} = 30.096.217,60 \\
 d_{90,3} &= \sqrt{\frac{165,86^2 - 799,93^2 + 91.353.782,40^2}{-185.500.072^2}} = 93.606.889,60 \\
 d_{91,1} &= \sqrt{\frac{165,86^2 - 117,44^2 + 98.253.072^2}{-65.786.287^2}} = 25.567.495,20 \\
 d_{91,2} &= \sqrt{\frac{165,86^2 - 220,05^2 + 98.253.072^2}{-121.989.400^2}} = 30.635.617,60 \\
 d_{91,3} &= \sqrt{\frac{165,86^2 - 799,93^2 + 98.253.072^2}{-185.500.072^2}} = 94.146.289,60 \\
 d_{92,1} &= \sqrt{\frac{154,33^2 - 117,44^2 + 96.587.472^2}{-65.786.287^2}} = 32.466.784,80 \\
 d_{92,2} &= \sqrt{\frac{154,33^2 - 220,05^2 + 96.587.472^2}{-121.989.400^2}} = 23.736.328 \\
 d_{92,3} &= \sqrt{\frac{154,33^2 - 799,93^2 + 96.587.472^2}{-185.500.072^2}} = 87.247.000
 \end{aligned}$$



$$\begin{aligned}
 d_{93,1} &= \sqrt{\frac{154,33^2 - 117,44^2 + 100.615.872^2}{-65.786.287^2}} = 30.801.184,80 \\
 d_{93,2} &= \sqrt{\frac{154,33^2 - 220,05^2 + 100.615.872^2}{-121.989.400^2}} = 25.401.928 \\
 d_{93,3} &= \sqrt{\frac{154,33^2 - 799,93^2 + 100.615.872^2}{-185.500.072^2}} = 88.912.600 \\
 d_{94,1} &= \sqrt{\frac{154,33^2 - 117,44^2 + 93.985.272^2}{-65.786.287^2}} = 34.829.584,80 \\
 d_{94,2} &= \sqrt{\frac{154,33^2 - 220,05^2 + 93.985.272^2}{-121.989.400^2}} = 21.373.528 \\
 d_{94,3} &= \sqrt{\frac{154,33^2 - 799,93^2 + 93.985.272^2}{-185.500.072^2}} = 84.884.200 \\
 d_{196,1} &= \sqrt{\frac{455,82^2 - 117,44^2 + 170.337.264,40^2}{-65.786.287^2}} = 104.550.977,20 \\
 d_{196,2} &= \sqrt{\frac{455,82^2 - 220,05^2 + 170.337.264,40^2}{-121.989.400^2}} = 48.347.864,40 \\
 d_{196,3} &= \sqrt{\frac{455,82^2 - 799,93^2 + 170.337.264,40^2}{185.500.072^2}} = 15.162.807,60 \\
 d_{197,1} &= \sqrt{\frac{455,82^2 - 117,44^2 + 167.780.184,40^2}{65.786.287^2}} = 101.993.897,20 \\
 d_{197,2} &= \sqrt{\frac{455,82^2 - 220,05^2 + 167.780.184,40^2}{121.989.400^2}} = 45.790.784,40 \\
 d_{197,3} &= \sqrt{\frac{455,82^2 - 799,93^2 + 167.780.184,40^2}{185.500.072^2}} = 17.719.887,60 \\
 d_{198,1} &= \sqrt{\frac{455,82^2 - 117,44^2 + 175.758.264,40^2}{-65.786.287^2}} = 109.971.977,20 \\
 d_{198,2} &= \sqrt{\frac{455,82^2 - 220,05^2 + 175.758.264,40^2}{-121.989.400^2}} = 53.768.864,40 \\
 d_{198,3} &= \sqrt{\frac{455,82^2 - 799,93^2 + 175.758.264,40^2}{-185.500.072^2}} = 9.741.807,61 \\
 d_{199,1} &= \sqrt{\frac{455,82^2 - 117,44^2 + 167.237.664,40^2}{-65.786.287^2}} = 101.451.377,20 \\
 d_{199,2} &= \sqrt{\frac{455,82^2 - 220,05^2 + 167.237.664,40^2}{121.989.400^2}} = 45.248.264,40 \\
 d_{199,3} &= \sqrt{\frac{455,82^2 - 799,93^2 + 167.237.664,40^2}{185.500.072^2}} = 18.262.407,60 \\
 d_{200,1} &= \sqrt{\frac{455,82^2 - 117,44^2 + 166.326.264,40^2}{65.786.287^2}} = 100.539.977,20 \\
 d_{200,2} &= \sqrt{\frac{455,82^2 - 220,05^2 + 166.326.264,40^2}{121.989.400^2}} = 44.336.864,40 \\
 d_{200,3} &= \sqrt{\frac{455,82^2 - 799,93^2 + 166.326.264,40^2}{185.500.072^2}} = 19.173.807,60
 \end{aligned}$$

Pada tabel 3 menunjukkan hasil perhitungan iterasi 1 kemudian memilih nilai terkecil dari setiap baris untuk menentukan cluster

Tabel 3. Hasil Iterasi Pertama

No	Semen (sak)	Biaya (Rp)	Cluster
1	223,57	73.119.168	1
2	223,57	71.749.848	1
3	223,57	75.085.368	1
4	223,57	69.543.768	1
5	223,57	69.097.368	1
...
90	165,86	91.353.782,40	1
91	154,33	98.253.072	2
92	154,33	96.587.472	2
93	154,33	100.615.872	2
94	154,33	93.985.272	2



No	Semen (sak)	Biaya (Rp)	Cluster
...
196	455,82	170.337.264,40	3
197	455,82	167.780.184,40	3
188	455,82	175.758.264,40	3
199	455,82	167.237.664,40	3
200	455,82	166.326.264,40	3
Jumlah Cluster 1:			72
Jumlah Cluster 2:			78
Jumlah Cluster 3:			50

Cluster Baru

Pembentukan cluster baru setelah dilakukan iterasi 1 kemudian dilakukan iterasi 2 untuk memastikan cluster yang dihitung tidak akan melakukan perubahan cluster yang ditunjukkan pada tabel 2 dengan nama kolom cluster, dimana tabel 3 kolom cluster pada iterasi 1 harus sama dengan kolom cluster pada iterasi 2 maka proses clustering selesai. Jika mengalami perubahan maka lakukan perhitungan cluster baru seperti yang ditunjukkan tabel 3. Berdasarkan iterasi 2 dan seterusnya sampai tidak ada perubahan cluster seperti yang ditunjukkan pada gambar 4.

$$C1 \text{ Semen} = \frac{223,57 + 223,57 + 223,57 + 223,57 + 223,57}{78} = 179,78$$

$$C1 \text{ Biaya} = \frac{73.119.168 + 71.749.848 + 75.085.368 + 69.543.768 + 69.097.368 + \dots + 91.353.782,40 + \dots}{78} = 73.194.977,37$$

$$C2 \text{ Semen} = \frac{\dots + 154,33 + 154,33 + 154,33 + 154,33}{72} = 304,53$$

$$C2 \text{ Biaya} = \frac{\dots + 98.253.072 + 96.587.472 + 100.615.872 + 93.985.272}{72} = 107.634.879,98$$

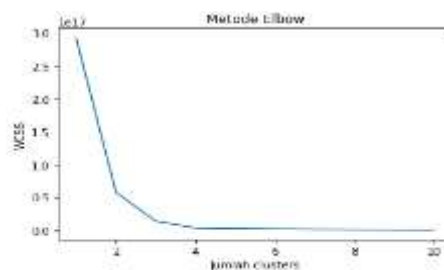
$$C3 \text{ Semen} = \frac{\dots + 455,82 + 455,82 + 455,82 + 455,82}{50} = 614,92$$

$$C3 \text{ Biaya} = \frac{\dots + 170.337.264,40 + 167.780.184,40 + 175.758.264,40 + 167.237.664,40 + 166.326.264,40}{50} = 172.112.780,20$$

Tabel 4. Cluster Baru

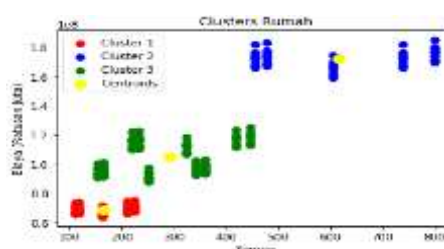
C1	179,78	73.194.977,37
C2	304,53	107.634.879,98
C3	614,92	172.112.780,20

Pada gambar 2 menunjukkan hasil jumlah cluster berdasarkan elbow dimana metode ini merupakan Teknik menetapkan berapa banyak cluster berdasarkan banyaknya cluster yang membentuk siku pada titik [17] [18]



Gambar 2. Menentukan Jumlah Cluster

Pada gambar 3 menunjukkan hasil visualisasi 200 data menggunakan K-Means dengan 3 cluster.



Gambar 3. Visualisasi Clustering K-Means



Gambar 4. Hasil Clustering K-Means

Pada gambar 4 menunjukkan hasil clustering dengan *K-Means* dimana *cluster* pertama sebanyak 50 data, *cluster* kedua sebanyak 50 data dan *cluster* ketiga sebanyak 100 data dengan waktu komputasi sebesar 0,444 detik dan *silhouette* 0,82

3.2 Penerapan Hierarchical

Dataset awal merupakan data dari pengembang perumahan

Pada tabel 5 merupakan potongan dataset *hierarchical* yang akan dihitung menggunakan persamaan (1).

Tabel 5. Dataset Hierarchical

No	Semen (Sak)	Biaya (Rp)
1	223,57	73.119.168
2	223,57	71.749.848
3	223,57	75.085.368
4	223,57	69.543.768
5	223,57	69.097.368
6	223,57	69.027.199,20
7	223,57	67.765.879,20
8	223,57	71.749.399,20
9	223,57	67.467.799,20
10	223,57	67.021.399,20
...
16	210,71	68.356.104
...
200	455,82	166.326.264,40

$$d_{2,1} = \sqrt{\frac{223,57^2 - 223,57^2 + 73.119.168^2 - 71.749.848^2}{2}} = 1.369.320$$

$$d_{3,2} = \sqrt{\frac{223,57^2 - 223,57^2 + 71.749.848^2 - 75.085.368^2}{2}} = 3.335.520$$

$$d_{4,3} = \sqrt{\frac{223,57^2 - 223,57^2 + 75.085.368^2 - 69.543.768^2}{2}} = 5.541.600$$

$$d_{5,4} = \sqrt{\frac{223,57^2 - 223,57^2 + 69.543.768^2 - 69.097.368^2}{2}} = 446.400$$

$$d_{6,5} = \sqrt{\frac{223,57^2 - 223,57^2 + 69.097.368^2 - 69.027.199,20^2}{2}} = 70.168,80$$

$$d_{7,6} = \sqrt{\frac{223,57^2 - 223,57^2 + 69.027.199,20^2 - 67.765.879,20^2}{2}} = 1.261.320$$

$$d_{8,7} = \sqrt{\frac{223,57^2 - 223,57^2 + 67.765.879,20^2 - 71.749.399,20^2}{2}} = 3.983.520$$

$$d_{9,8} = \sqrt{\frac{223,57^2 - 223,57^2 + 71.749.399,20^2 - 67.467.799,20^2}{2}} = 3.983.520$$

$$d_{10,9} = \sqrt{\frac{223,57^2 - 223,57^2 + 67.467.799,20^2 - 67.021.399,20^2}{2}} = 446.400$$

Gunakan persamaan (1) untuk menghitung dari data ke 16, 17 sampai data ke 200

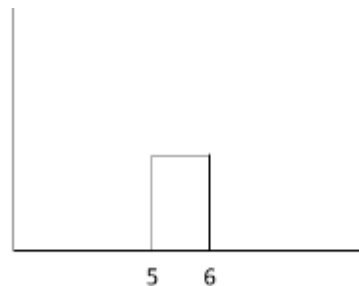
Tabel 6. Hasil Single Linkage Pertama

Dat a	1	2	3	4	5	6	7	8	9	10	..	20 0
1	0											



Dat a	1	2	3	4	5	6	7	8	9	10	..	20
2	1.369.32 0	0										0
3	3.335.52 0	3.335.52 0	0									
4	5.541.60 0	5.541.60 0	5.541.60 0	0								
5	446.400	446.400	446.400	446.400	0							
6	70.168,8 0	70.168,8 0	70.168,8 0	70.168,8 0	70.168,8 0	0						
7	1.261.32 0	1.261.32 0	1.261.32 0	1.261.32 0	1.261.32 0	1.261.32 0	0					
8	3.983.52 0	3.983.52 0	3.983.52 0	3.983.52 0	3.983.52 0	3.983.52 0	3.983.52 0	0				
9	3.983.52 0	3.983.52 0	3.983.52 0	3.983.52 0	3.983.52 0	3.983.52 0	3.983.52 0	3.983.52 0	0			
10	446.400	446.400	446.400	446.400	446.400	446.400	446.400	446.400	446.400	0		
...	0	
16	70.168,8 0	70.168,8 0	70.168,8 0	70.168,8 0	70.168,8 0	70.168,8 0	70.168,8 0	70.168,8 0	70.168,8 0	70.168,8 0	...	
...	
200	911.400	911.400	911.400	911.400	911.400	911.400	911.400	911.400	911.400	911.400	...	0

Pada tabel 6 terdapat warna kuning merupakan hasil perhitungan dengan memilih nilai terkecil pada matrik 5,6 sehingga hasil dendogramnya ditunjukkan pada gambar 5



Gambar 5. Hasil Dendogram Pertama

Tabel 7. Hasil Single Linkage Kedua

Dat a	(6,5)	2	3	4	7	8	9	10	...	15	1	...	20
(6,5)	0										6	...	0
2	70.168,8 0	0											
3	70.168,8 0	3.335.52 0	0										
4	70.168,8 0	5.541.60 0	5.541.60 0	0									
7	1.261.32 0	1.261.32 0	1.261.32 0	1.261.32 0	0								
8	3.983.52 0	3.983.52 0	3.983.52 0	3.983.52 0	3.983.52 0	0							
9	3.983.52 0	3.983.52 0	3.983.52 0	3.983.52 0	3.983.52 0	3.983.52 0	0						
10	446.400	446.400	446.400	446.400	446.400	446.400	446.400	0					
...	0				
16	70.168,8 0	70.168,8 0	70.168,8 0	70.168,8 0	70.168,8 0	70.168,8 0	70.168,8 0	70.168,8 0	...	70.168,8 0	0	...	
...	
200	911.400	911.400	911.400	911.400	911.400	911.400	911.400	911.400	0

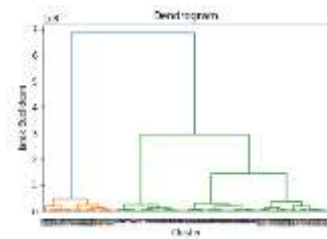
Pada tabel 7 terdapat warna kuning merupakan hasil perhitungan dengan memilih nilai terkecil pada matrik 15,16 sehingga hasil dendogramnya ditunjukkan pada gambar 6



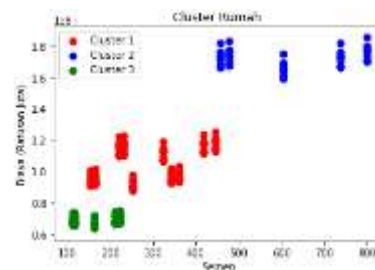
Gambar 6. Hasil Dendogram Kedua



Berdasarkan hasil tabel 7 maka dapat digambarkan *dendrogram* yang ditunjukkan pada gambar 7. Lakukan seperti tabel 6 dan tabel 7 sampai data ke 200 berdasarkan data pada tabel 4, sehingga akan seperti pada gambar 8.

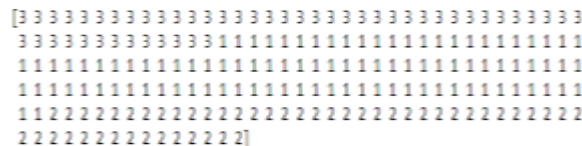


Gambar 7. Dendrogram Keseluruhan



Gambar 8. Visualisasi Clustering Agglomerative Hierarchical

Pada gambar 8 menunjukkan hasil visualisasi 200 data menggunakan *Agglomerative Hierarchical* dengan 3 cluster.



Gambar 9. Hasil Clustering Agglomerative Hierarchical

Pada gambar 9 menunjukkan hasil *clustering* dengan *Agglomerative Hierarchical* dimana *cluster* pertama sebanyak 100, *cluster* kedua sebanyak 50 dan *cluster* ketiga sebanyak 50 dengan waktu komputasi sebesar 3,22 detik dan *silhouette* 0,51

4. KESIMPULAN

Berdasarkan hasil dari clustering dengan menggunakan metode *K-Means* dan *Agglomerative Hierarchical* menggunakan *single linkage* maka dapat disimpulkan bahwa kelompok dua menghasilkan data yang sama tetapi untuk kelompok satu terdapat perbedaan hasil dimana metode *K-Means* menghasilkan hasil *cluster* dengan data lebih kecil sedangkan untuk kelompok 1 *K-Means* menghasilkan hasil *cluster* lebih besar. Waktu yang dibutuhkan dari proses perhitungan, visualisasi data *K-Means* lebih cepat dari *Agglomerative Hierarchical* serta nilai *silhouette* pada *K-Means* lebih baik dari *Agglomerative Hierarchical*. Saran untuk penelitian selanjutnya dapat mencoba menggunakan metode *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) untuk pengelompokan datanya

REFERENCES

- [1] D. I. Pratama, "Analisis Anggaran Pelaksanaan Pembangunan Rumah Tinggal (Studi Kasus: Rumah Tipe 50/97 di Perumahan Dian Arta – Bangunjiwo, Bantul)," 2018.
- [2] V. L. Deanggi, A. S. B. Nugroho, and F. Siswanto, "Pengaruh Biaya Prasarana, Sarana, Utilitas Dan Overhead Terhadap Biaya Produksi Unit Bangunan Pada Perumahan Mewah Di Yogyakarta," *INERSIA INformasi dan Ekspose Has. Ris. Tek. Sipil dan Arsit.*, vol. 15, no. 2, pp. 72–83, 2019.
- [3] A. Halim, "Perbandingan kuat tekan dan kuat geser spesi tembok yang digunakan masyarakat," vol. 20, no. 1, pp. 1–5, 2012.
- [4] K. Khomsatun, D. Ikhsan, M. Ali, and K. Kursini, "Sistem Pengambilan Keputusan Pemilihan Lahan Tanam Di Kabupaten Wonosobo Dengan K-Means Clustering Dan Topsis," *J. Nas. Pendidik. Tek. Inform.*, vol. 9, no. 1, p. 55, 2020.
- [5] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019.
- [6] 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.
- [7] W. Afifi, D. R. Nastiti, and Q. Aini, “Clustering K-Means Pada Data Ekspor (Studi Kasus: Pt. Gaikindo),” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 1, pp. 45–50, 2020.
- [8] G. Abdurrahman, “Clustering Data Kredit Bank Menggunakan Algoritma Agglomerative Hierarchical Clustering Average Linkage,” *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones.)*, vol. 4, no. 1, p. 13, 2019.
- [9] I. Rahma, P. Prima Arhandi, and A. Tufika Firdausi, “Penerapan Metode Hierarchical Clustering Dan K-Means Clustering Untuk Mengelompokkan Potensi Lokasi Penjualan Linkaja,” *J. Inform. Polinema*, vol. 6, no. 1, pp. 15–22, 2020.
- [10] W. Widyawati, W. L. Y. Saptomo, and Y. R. W. Utami, “Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan,” *J. Ilm. SINUS*, vol. 18, no. 1, p. 75, 2020.
- [11] E. Suherman, “Agglomerative Hierarchical Clustering Dengan Berbagai Pengukuran Jarak Dalam Mengklaster Daerah Berdasarkan Tingkat Kemiskinan,” vol. 5, no. 1, pp. 978–979, 2019.
- [12] Emi Ariska, “Implementasi Agglomerative Hierarchical Clustering Pada Data Produksi Dan Data Penjualan Perusahaan,” Universitas Sumatera Utara, 2018.
- [13] “Perbedaan Supervised Learning and Unsupervised Learning.” [Online]. Available: <https://www.uc.ac.id/ict/perbedaan-supervised-learning-and-unsupervised-learning/>.
- [14] R. Bonnin, *Building machine learning projects with TensorFlow: engaging projects that will teach you how complex data can be exploited to gain the most insight*. 2016.
- [15] S. F. Mu’afa and N. Ulinuha, “Perbandingan Metode Single Linkage, Complete Linkage Dan Average Linkage dalam Pengelompokan Kecamatan Berdasarkan Variabel Jenis Ternak Kabupaten ...,” ... *J. Ilm. Bid. Teknol. ...*, vol. 4, no. 2, 2019.
- [16] A. Aditya, I. Jovian, and B. N. Sari, “Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019,” *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 51, 2020.
- [17] D. A. I. C. Dewi and D. A. K. Pramita, “Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali,” *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019.
- [18] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, “Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, no. 1, 2018.