ICIT Journal

ISSN : 2356 – 5195

Online ISSN: 2654 - 8704

Perbandingan Manhattan dan Euclidean Distance Untuk Pengelompokan Penyakit Jantung Menggunakan Algoritma K-Means

Nicholas Febrian Sutirta[1], Noviandi*[2]

[1,2]Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Esa Unggul

E-mail: [1]nicholasf249@student.esaunggul.ac.id, *[2]noviandi@esaunggul.ac.id

Abstrak

Penelitian ini menggunakan metode perbandingan jarak antara Manhattan distance dan Euclidean distance dalam menentukan keakuratan k-means. Fokus penelitian ini adalah untuk menentukan perbandingan keakuratan antara dua jarak yang digunakan untuk metode k-means clustering. Dengan menggunakan metode k-means penelitian ini bertujuan untuk mengelompokan cluster penyakit jantung, dengan menggunakan dataset yang memiliki atribut kolesterol dan umur, untuk menentukan keakuratan dari kedua pengukuran jarak k-means penulis menggunakan silhouette coeffiecient. Hasil dari penelitian ini menunjukan bahwa pemilihan pengukuran jarak untuk k-means clustering mempengaruhi hasil akhir dari clustering dan silhouette score yaitu 0.5374 untuk silhouette score dari k-means yang menggunakan Manhattan distance dan 0.5355 untuk silhouette score dari k-means yang menggunakan Euclidean distance. Hasil tersebut menunjukan bahwa k-means menggunakan Manhattan distance menghasilkan silhouette score yang lebih tinggi dari Euclidean distance. Penelitan lebih lanjut tentang pengaruh dari jarak perbandingan untuk k-means clustering dengan menggunakan dataset berbeda dapat memberikan wawasan lebih lanjut yang berguna dalam layanan kesehatan dan bidang lainya.

Kata Kunci: K-means clustering, Jarak Euclidean, Jarak Manhattan, Sillhouette coefficient.

Abstract

This research use the comparison method between Manhattan distance and Euclidean distance to determine the accuracy of k-means. The objective of this study is to determine the accuracy between the two distances used for the method of k-means clustering. Using k-means clustering, this research goal is to cluster heart diseases using datasets of cholesterol and age attributes to determine the accuracy of Manhattan and Euclidean k-means distance metrics. To determine the accuracy of the two distances, the authors of this study use the silhouette coefficient method. The result section of this study indicates that the choice of the distance metrics used in k-means clustering affects the plot and silhouette scores. K-means clustering that uses Manhattan distance yields a silhouette score of 0.5374, and k-means that use Euclidean distance as its distance metrics have a silhouette score of 0.5355. Therefore, the results show that k-means clustering that uses Manhattan distance as its distance metric gives a slightly higher silhouette score compared to the k-means that use Euclidean distance. Further research on the comparison of k-means distances using different datasets could provide useful insights in healthcare related domains.

Keywords: *K-means clustering, Euclidean distance, Manhattan distance, Sillhouette coefficient.*

ISSN : 2356 – 5195
Online ISSN: 2654 - 8704

1. PENDAHULUAN

Penyakit jantung adalah salah satu penyakit utama penyebab kematian terbanyak di dunia, diperkirakan sekitar 17,9 juta orang penderita penyakit kardiovaskular 85% diantaranya adalah penderita penyakit jantung[1]. Hasil dari Riset Kesehatan Dasar (Riskesdas) tahun 2018 menunjukan prevalensi penyakit jantung sebesar 1,5% dari seluruh sampel penduduk Indonesia[2]. Penyebab utama penyakit jantung adalah pola hidup yang kurang sehat seperti makanan, jarang olahraga, merokok dan alcohol[3]. Efek dari pola hidup tersebut adalah kenaikan tekanan darah, gula darah dan obesitas, gejala gejala ini adalah faktor utama pemicu terjadinya penyakit jantung[4]. Beberapa tipe penyakit jantung adalah penyakit jantung coroner, angina, serangan jantung, gagal jantung, aritmia, katup jantung, dan penyakit jantung keturunan [5].

Terdapat beberapa penelitian untuk menganalisis penyakit jantung menggunakan Algoritma k-means antara lain adalah jurnal yang dibuat Trisca Anggraini Putri, dkk [6]. Selain itu pada jurnal yang ditulis oleh Al Rivan dan R. A. Sonaru juga menggunakan algoritma k-means untuk dijadikan perbandingan, jurnal tersebut menggunakan Euclidean Distance untuk menghitung jarak antar titik pusat dan titik objek[7].

Karena algoritma k-means adalah algoritma yang cukup sederhana, mudah digunakan, dan sangat efisien saat memproses data berskala besar, algoritma k-means sering digunakan banyak orang. Algoritma ini memiliki alur yang sederhana, dengan memasukan data dan kemudian dikelompokan dengan nilai k. Namun ada beberapa kelemahan dalam algoritma ini, contohnya dalam menggunakan Euclidean distance tingkat perbedaan dalam pengelompokan data sangat rendah sehingga nilai yang dihasilkan tidak menentu [8].

Dari beberapa penelitian terdahulu, belum ada yang menggunakan manhattan distance untuk membandingkan data penyakit jantung yang dihasilkan algoritma k-means. Oleh karena itu pada penelitian ini peneliti menggunakan perbandingan Manhattan distance dan Euclidean distance untuk pengelompokan penyakit jantung dengan tujuan untuk mengetahui ke-akuratan dari kedua metode tersebut. Data yang didapat dari kaggle nantinya akan dianalisis menggunakan algoritma K-means dan dibandingkan menggunakan Manhattan distance dan Euclidean distance, hasilnya adalah bentuk perbandingan dari kedua analisis tersebut.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Dataset yang akan digunakan dalam penelitian ini berupa record data dari kaggle yang berjudul “Heart Disease”. Dataset ini memiliki 1025 record data dengan berbagai macam atribut yang akan digunakan pada penelitian ini.

Dataset ini diambil dari tahun 1988 yang dikumpulkan dari empat sumber database (Cleveland, Hungaria, Swiss, dan Long Beach V). Data ini berisi 76 atribut, dan dari 76 data yang dipakai hanya 13 atribut yaitu umur, jenis kelamin, jenis rasa sakit dalam dada, tekanan darah biasa, kolesterol, gula darah puasa, electrocardio graphic, detak jantung maksimal, angina, ST karena istirahat, pembuluh darah yang membengkak, dan kelainan jantung.

2.2 *Preprocessing* Data

Preprocessing data adalah proses untuk mengubah data mentah menjadi data yang siap untuk digunakan pada penelitian ini, preprocessing data dilakukan dengan tujuan agar data yang akan diolah dapat dianalisis dengan baik[9].

Proses yang dilakukan di tahapan ini adalah mengidentifikasi dan mengisi atau menghapus nilai yang hilang dalam kumpulan data.

ISSN : 2356 – 5195
Online ISSN: 2654 - 8704

```
Data columns (total 13 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       1025 non-null   int64
 1   sex       1025 non-null   int64
 2   cp        1025 non-null   int64
 3   trestbps  1025 non-null   int64
 4   chol      1025 non-null   int64
 5   fbs       1025 non-null   int64
 6   restecg   1025 non-null   int64
 7   thalach   1025 non-null   int64
 8   exang     1025 non-null   int64
 9   oldpeak   1025 non-null   float64
 10  slope     1025 non-null   int64
 11  ca        1025 non-null   int64
 12  thal      1025 non-null   int64
dtypes: float64(1), int64(12)
```

Gambar 2.1 Info data yang digunakan

Data yang akan digunakan menggunakan data dalam bentuk angka (numerik) sehingga mudah untuk dianalisis dari 1025 record data.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	2
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	2
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3

1025 rows × 13 columns

Gambar 2.2 Dataset yang digunakan untuk penelitian

2.3 *Clustering*

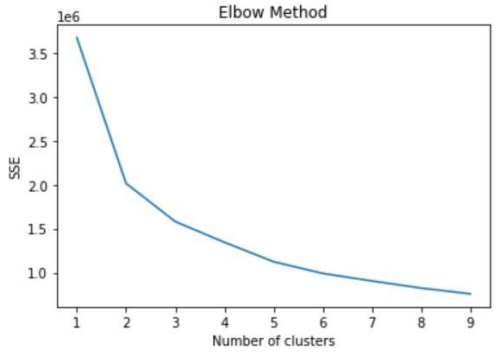
Pada proses clustering tahapan pertama yang akan dilakukan adalah menentukan nilai k yang optimal untuk digunakan dalam algoritma k-means, kemudian setelah algoritma k-means, dilakukan penghitungan jarak menggunakan Euclidean dan Manhattan distance, dan evaluasi keakuratan perbandingan antara Euclidean distance dan Manhattan distance.

2.3.1 Penentuan Nilai k

Pada penelitian ini penentuan nilai k ditentukan menggunakan Elbow Method. Elbow method ini dihitung berdasarkan kuadrat jarak antar titik sampel di pusat cluster dan pusat cluster untuk memberikan serangkaian nilai k, nilai dari k ditentukan dengan menentukan kurva yang melengkung signifikan kebawah, dititik tersebut menunjukan nilai k yang paling optimal[10].

Di proses elbow method ini pertama kita akan membuat daftar kosong yang disebut sse(sum of squared errors) dibuat untuk menyimpan nilai SSE untuk setiap jumlah cluster. Kemudian membuat loop yang menjalankan algoritma k-means untuk nilai k 1 sampai 10, berikut ini adalah hasil dari elbow method yang digunakan untuk menentukan nilai k.

ISSN : 2356 – 5195
Online ISSN: 2654 - 8704



Gambar 2.3 Elbow Method untuk menentukan jumlah k

Dari elbow method yang dihasilkan, kita dapat menentukan jumlah k yang optimal untuk digunakan pada penelitian ini, digambar tersebut memperlihatkan pada grafis antara 1 dan 2 menunjukan penurunan yang signifikan dan dari 2 ke 3 menunjukan penurunan yang tidak signifikan, dari gambar tersebut dapat kita simpulkan bahwa nilai k yang optimal untuk digunakan untuk dataset ini adalah 2.

Table 2.1 Nilai SSE menggunakan Elbow Method

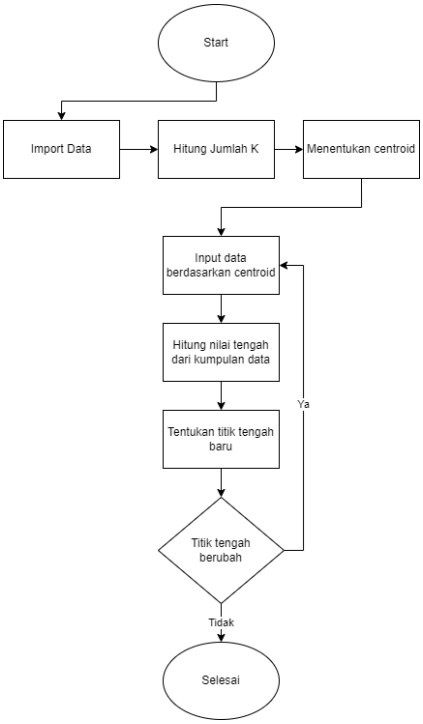
No	SSE	Difference
1	3671607.176	0
2	2014840.852	1656766.32
3	1580652.498	434188.354
4	1344984.402	235668.096
5	1123983.06	221001.342
6	990795.152	133187.908
7	904984.0417	85811.1103
8	824073.304	80910.7377
9	758895.9636	65177.3404
10	698946.3002	59949.6634

2.3.2 Algoritma K-Means Menggunakan *Euclidean Distance*

Tahapan yang akan dilakukan untuk menerapkan algoritma k-means untuk mengelompokan data adalah sebagai berikut:

1. Mengimport dataset yang akan dikelompokan
2. Menentukan jumlah k
2. Menentukan centroid secara random berdasarkan nilai k
3. Mengelompokan setiap data yang memiliki kemiripan berdasarkan centroid yang ditentukan
4. Input setiap data
5. Tentukan nilai centroid yang baru
6. Apabila semua data sudah dikelompokan dan nilai centroid tidak berubah, selesai

ISSN : 2356 – 5195
Online ISSN: 2654 - 8704



Gambar 2.4 Flowchart K-means

Euclidean distance adalah metode pengukuran jarak yang mengacu pada jarak antara dua titik di suatu dimensi[8]. Jarak Euclidean distance dapat ditentukan dengan menggunakan persamaan dibawah ini:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad \textbf{(1)}$$

Algoritma k-means menggunakan Euclidean distance sebagai metode pengukuran jarak, oleh karena itu perbandingan antara jarak pengelompokan cluster menggunakan Euclidean distance dapat dihitung dengan menghitung jarak antar centroid yang dihasilkan oleh algoritma k-means.

2.3.3 Algoritma K-Means Menggunakan *Manhattan Distance*

Manhattan distance adalah metode pengkuran jarak yang digunakan untuk menghitung jarak mutlak antara dua koordinat suatu objek, metode pengukuran jarak ini juga biasa disebut cityblock distance atau taxicab metrics[11].

Perhitungan jarak dengan menggunakan Manhattan distance menggunakan dua titik koordinat objek. Rumus yang digunakan dalam perhitungan Manhattan distance untuk membandingkan kedua jarak perhitungan yang digunakan adalah sebagai berikut:

$$d = \sum_{i=1}^{n} |x_i - y_i| \quad \textbf{(2)}$$

Dengan menggunakan persamaan tersebut, kita dapat mengetahui perbandingan kedua jarak data yang ingin dihitung.

2.4 Evaluasi

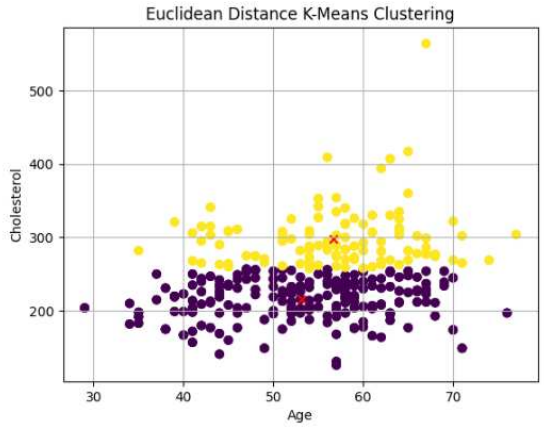
Data yang sudah di analisis dengan algoritma k-means menggunakan metode

perbandingan jarak manhattan distance akan dibandingkan keakuratanya dengan Euclidean distance. Hasilnya adalah visualisasi pengelompokan data penyakit jantung antara perbandingan jarak Manhattan distance dengan Euclidean distance.

3. HASIL DAN PEMBAHASAN

3.1 Model Clustering Dataset

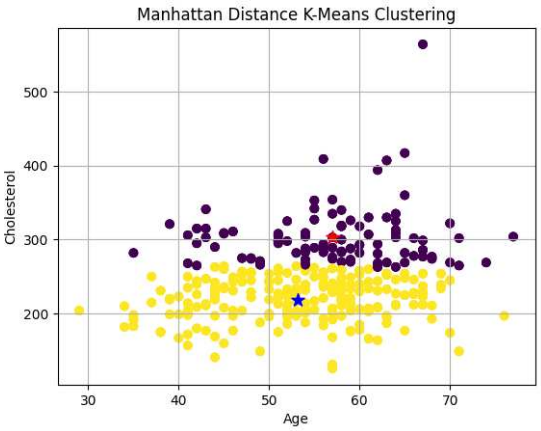
Setelah dilakukan preprocessing data dan penentuan nilai k menggunakan metode elbow method yang sudah dilakukan pada dataset 'heart disease' ini memiliki nilai optimal k=2 maka proses clustering yang akan dilakukan akan menggunakan nilai k=2. Algoritma k-means menggunakan Euclidean distance dengan menggunakan persamaan **2-1** dan nilai k=2 menghasilkan grafik seperti berikut ini:



Gambar 3.1 K-means Clustering Menggunakan Euclidean Distance

Berdasarkan graphic di atas data dibagi menjadi 2 cluster yaitu cluster dengan kolesterol rendah dan cluster dengan kolesterol tinggi. Cluster dengan kolesterol rendah berkisar dari nilai 126 sampai 257 dan cluster dengan kolesterol tinggi dengan nilai dari 258 sampai 564.

Sedangkan Manhattan distance menggunakan persamaan **2-3** dan menggunakan nilai k=2 menghasilkan plot seperti berikut ini:



Gambar 3.2 K-means Clustering Menggunakan Manhattan Distance

K-means clustering menggunakan manhattan distance memiliki 2 cluster yang sama dengan k-means clustering menggunakan Euclidean distance yaitu memiliki cluster pertama dengan nilai terendah di nilai kolesterol 263 dan nilai tertinggi di 564 dan cluster 2 dengan nilai kolesterol terendah di 126 dan tertinggi di 264.

3.2 Pembahasan Hasil Penelitian

Hasil dari cluster k-means dengan Euclidean distance menunjukan tinggi rendahnya kolesterol berdasarkan umur dari data yang sudah dikumpulkan. Data pasien yang memiliki kolesterol rendah berada pada cluster 1 yang memiliki rata kolesterol di 215 mg/dL, sedangkan cluster 2 termasuk data dengan tingkat kolesterol yang tinggi dengan rata-rata kolesterol 298 mg/dL.

Kemungkinan untuk pasien yang ada di cluster 1 atau memiliki kolesterol rendah untuk terkena penyakit jantung lebih kecil, sedangkan pasien yang ada di cluster 2 memiliki kemungkinan lebih tinggi untuk terkena penyakit jantung. Dalam data yang digunakan untuk dianalisis pasien yang memiliki kadar kolesterol tergolong baik berkisar antara 126 mg/dL sampai 257mg/dL, sedangkan pasien yang tergolong memiliki kolesterol cukup tinggi memiliki nilai antara 258 mg/dL sampai 564 mg/dL. Kolesterol bukan penentu apakah seseorang memiliki penyakit jantung atau tidak, karena penyakit jantung juga memiliki atribut lain penyebab penyakit jantung seperti rasa sakit di dada, tekanan darah, gula darah, namun kolesterol merupakan salah satu atribut utama dalam menentukan seseorang beresiko memiliki penyakit jantung atau tidak.

Table 3.1 Hasil cluster euclidean distance

Euclidean distance			
Cluster	Age	Chol	Total data
1	53.12	215.5	648
2	56.7	298.43	377

Berdasarkan pola yang dihasilkan algoritma k-means clustering dengan Euclidean distance, hasil dari analisis cluster 1 memiliki 648 pasien dengan tingkat kolesterol rendah yang memiliki kemungkinan lebih rendah memiliki penyakit jantung, sedangkan cluster 2 memiliki 377 pasien dengan tingkat kolesterol tinggi yang memiliki kemungkinan lebih tinggi terkena penyakit jantung.

Table 3.2 Hasil cluster manhattan distance

Manhattan distance			
Cluster	Age	Chol	Total data
1	57.02	302.79	338
2	53.16	218.06	687

Sedangkan untuk Manhattan distance menunjukan hasil yang tidak jauh berbeda dengan Euclidean distance. Pada cluster 1 kmeans Manhattan distance memiliki 338 pasien dengan tingkat kolesterol tergolong tinggi dan cluster 2 memiliki data pasien 687 dengan tingkat kolesterol rendah. Cluster 1 memiliki pasien dengan rata rata kolesterol 302mg/dL sedangkan cluster 2 memiliki pasien dengan rata rata kolesterol 218 mg/dL.

3.3 Analisis Data

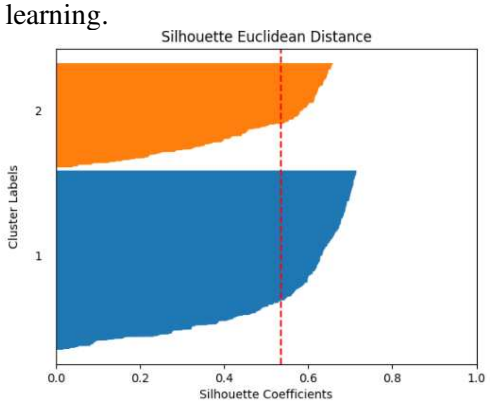
Klasifikasi tinggi rendahnya kolesterol seseorang berdasarkan total kolesterol adalah dibawah 200mg/dL untuk batas wajar atau normal, 200mg/dL sampai 239mg/dL untuk dapat dikatakan tinggi untuk penderita penyakit jantung, dan diatas 240mg/dL adalah tingkat kolesterol tinggi dan memiliki resiko dua kali lipat lebih tinggi terkena penyakit jantung dibandingkan orang orang yang memiliki tingkat kadar total kolesterol dibawah rentang tersebut[12]. Namun tingkat kolesterol total juga dipengaruhi oleh kolesterol HDL dan kolesterol LDL. Rentang kolesterol

ISSN : 2356 – 5195
Online ISSN: 2654 - 8704

LDL seseorang bervariasi berdasarkan resiko dan factor factor lain yang dimiliki orang tersebut. Apabila seseorang memiliki penyakit jantung atau diabetes orang tersebut memiliki rentang LDL yang lebih rendah dibandingkan orang yang tidak memiliki penyakit jantung atau resiko penyakit lainnya.

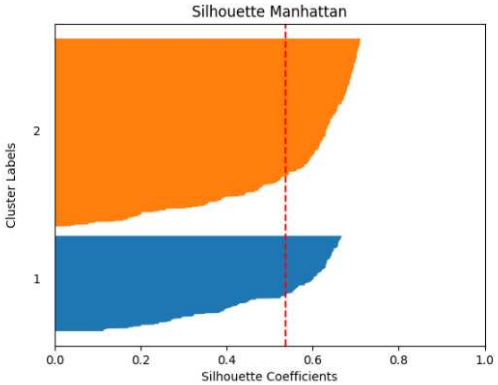
Berdasarkan penilitian yang dilakukan, data yang sudah dikelompokan menunjukan rata rata kolesterol 215mg/dL dalam kelompok cluster rendah dan memiliki rata rata kolesterol 298mg/dL untuk pengelompokan yang dilakukan menggunakan Euclidean distance. Sedangkan untuk manhattan distance cluster dengan kolesterol rendah menunjukan 218mg/dL dan cluster dengan kolesterol tinggi menunjukan 302mg/dL. Hasil penelitian tersebut menunjukan hasil yang cukup bagus jika dibandingkan dengan jurnal peneliti sebelumnya berdasarkan tinggi rendah nya total kolesterol dan hubunganya dengan penyakit jantung.

Penelitian ini menggunakan silhouette coefficient untuk menentukan keakuratan atau score dari kedua perbandingan jarak yang digunakan algoritma kmeans, berdasarkan jurnal [13], silhouette coefficient memiliki hasil yang cukup memuaskan dalam mengevaluasi model clustering dalam machine learning.



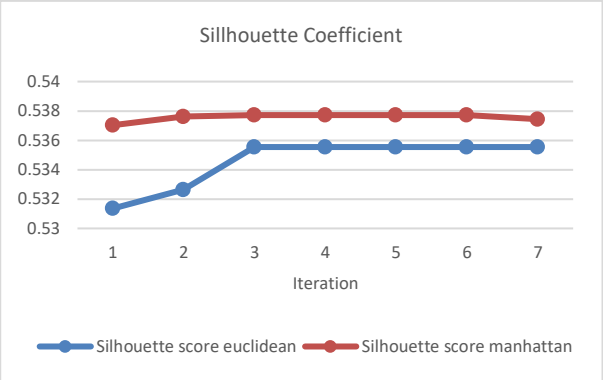
Gambar 3.3 Hasil silhouette coefficient euclidean distance

Dari data yang sudah dikelompokan menggunakan nilai iterasi maksimal yaitu 7, hasil evaluasi menggunakan silhouette coefficient menunjukan score 0.5355 untuk kmeans dengan Euclidean distance sebagai metode perhitungan jaraknya.



Gambar 3.4 Hasil silhouette coefficient manhattan distance

Sedangkan hasil untuk k-means menggunakan metode manhattan distance dengan iterasi maksimal yaitu 7 menunjukan silhouette coefficient dengan nilai 0.5374.



Gambar 3.5 Sillhouette coefficient

Dari 7 kali iterasi yang dilakukan dalam proses clustering ini, didapatkan perbandingan iterasi yang memiliki nilai lebih tinggi pada iterasi awal dibandingkan iterasi terakhir. Kedua perbandingan silhouette coefficient menunjukan nilai yang tidak signifikan untuk manhattan distance dan Euclidean distance.

Nilai silhouette coefficient memiliki rentang dari -1 sampai 1 dimana apabila nilai silhouette coefficient semakin mendekati 1 berarti data sampel memiliki jarak yang cukup jauh dari kelompok lainnya dan nilai dibawah 0 berarti data sampel memiliki jarak mendekati atau mengenai cluster lainya, dengan nilai 0.52 – 0.70 menandakan struktur clustering yang cukup baik[14]. Berdasarkan nilai dari perbandingan silhouette coefficient diatas dapat disimpulkan bahwa manhattan distance memiliki score yang lebih baik dari manhattan distance untuk kmeans clustering menggunakan dataset heart disease ini.

4. KESIMPULAN

Berdasarkan pengujian dan perancangan yang dilakukan dalam penelitian ini, maka dapat disimpulkan bahwa keakuratan model k-means dengan menggunakan metode Manhattan distance dengan 7 kali iterasi berdasarkan metode silhouette coefficient adalah 0.5374. keakuratan model kmeans dengan menggunakan metode Euclidean distance dengan 7 kali iterasi berdasarkan metode silhouette coefficient adalah 0.5355. Berdasarkan hasil silhouette coefficient yang dihasilkan kedua model kmeans dengan metode Manhattan dan Euclidean distance dapat disimpulkan bahwa metode manhattan distance memiliki nilai yang lebih baik dengan selisih nilai 0.0019.]

5. SARAN

Setelah disimpulkan bahwa penelitian ini masih memiliki banyak kekurangan, penulis mengharapkan penelitian selanjutnya dapat menghilangkan atau mengurangi berbagai kekurangan yang ada pada penelitian ini. Saran untuk penelitian selanjutnya yang akan membandingkan keakuratan kedua metode perbandingan jarak adalah untuk menggunakan data yang digunakan lebih bervariasi agar memiliki nilai keakuratan yang lebih baik atau menggunakan metode lain dengan tingkat score keakuratan yang lebih tinggi lagi.

DAFTAR PUSTAKA

[1] M. B. Liu, "Cardiovascular diseases," *Chinese Medical Journal*, 2014. https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed Nov. 22, 2022).
[2] Kemenkes RI, "Hasil Riset Kesehatan Dasar Tahun 2018," *Kementrian Kesehat. RI*, vol. 53, no. 9, pp. 1689–1699, 2018.
[3] Centers for Disease Control and Prevention (CDC), "About Heart Disease & Stroke," *Million Hearts*, 2012. http://millionhearts.hhs.gov/abouthds/cost-consequences.html (accessed Nov. 22, 2022).
[4] K. J. Ho, "Cardiovascular diseases," *Nutritional Aspects of Aging: Volume 2*, 2018. https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1 (accessed Nov. 22, 2022).
[5] Fabiana Meijon Fadul, "Common heart condition," 2013. https://www.nhsinform.scot/illnesses-and-conditions/heart-and-blood-vessels/conditions/common-heart-conditions (accessed Dec. 20, 2022).
[6] T. A. Putri and N. Huda, "Analisis dan Prediksi Penyakit Jantung Menggunakan Algoritma K-Means Clustering Pada Rumah Sakit Umum," *Bina Darma Conf. Comput. Sci.*, vol. 4, no. 41, pp. 197–206, 2020.
[7] M. E. Al Rivan and R. A. Sonaru, "Perbandingan Metode K-Means dan GA K-Means untuk Clustering Dataset Heart Disease Patients," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2585–2597, 2022, doi: 10.35957/jatisi.v9i3.2799.
[8] A. Zhu, Z. Hua, Y. Shi, Y. Tang, and L. Miao, "An improved k-means algorithm based on evidence distance," *Entropy*, vol. 23, no. 11, 2021, doi: 10.3390/e23111550.
[9] N. Noviandi, S. A. Noviantika, and B. Irawan, "Clustering Villages Based on Distance and Accessibility to Health Facilities Using the K-Means Method," *J. Teknol. Dan Open Source*, vol. 5, no. 1, pp. 35–42, 2022, doi: 10.36378/jtos.v5i1.2184.
[10] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J*, vol. 2, no. 2, pp. 226–235, 2019, doi: 10.3390/j2020016.
[11] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 20–24, 2019, doi: 10.30591/jpit.v4i1.1253.
[12] M. Hongbao, "Cholesterol and Human Health," *Nat. Sci.*, vol. 2, no. 4, pp. 17–21, 2004.
[13] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, pp. 1–17, 2021, doi: 10.3390/e23060759.
[14] B. N. Sari, "Identification of Tuberculosis Patient Characteristics Using K-Means Clustering," *Sci. J. Informatics*, vol. 3, no. 2, pp. 129–138, 2016, doi: 10.15294/sji.v3i2.7909.