

Implementasi Metode K-NN dalam Klasterisasi Kasus Kesehatan Jantung

Anggraini Puspita Sari*, Andreas Nugroho Sihananto, Dwi Arman Prasetya

Program Studi Sains Data, Universitas Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia

*anggraini_ps8@gmail.com

Kata Kunci :

ABSTRAK

*K-NN
Clustering
Kesehatan jantung*

Penyakit jantung penyebab kematian nomor satu berdasarkan data yang diperoleh dari WHO (world health organization). Penyakit jantung terjadi ketika darah yang mengalir ke otot jantung berhenti sehingga menyebabkan gangguan jantung. Hal ini menyebabkan adanya kebutuhan mendefinisikan sistem pendukung keputusan yang membantu dokter dalam mengambil keputusan untuk mengambil tindakan pencegahan terhadap penderita penyakit jantung. K-NN (K-Nearest Neighbor) merupakan metode yang sangat sederhana, paling populer, sangat efisien dan efektif untuk pengenalan pola. K-NN merupakan pengklasifikasi lurus ke depan dengan sampel diklasifikasikan berdasarkan kelas tetangga terdekatnya. Basis data medis memiliki volume tinggi. Jika kumpulan data berisi atribut yang berlebihan dan tidak relevan, maka klasifikasi dapat menghasilkan hasil yang kurang akurat. Penelitian ini menerapkan metode klasifikasi K-NN diharapkan dapat mengatasi permasalahan untuk efektifitas dan akurasi dalam mendeteksi kesehatan jantung. Dalam penelitian ini mencakup pengukuran performa, yaitu: presisi, *recall*, *f-measure*, dan akurasi menggunakan metode K-NN dengan nilai $K = 3$. Dataset yang digunakan dari *UCI Machine Learning Repository* pada 303 pasien penyakit jantung. Hasil yang didapatkan ialah presisi 0.70, *recall* 0.94, dan *f-measure* 0.81, dan akurasi 70% yang termasuk dalam klasifikasi baik dari nilai K terdekat sehingga metode K-NN dapat digunakan dalam mendeteksi kesehatan jantung.

1. Pendahuluan

Penyakit Jantung merupakan kondisi penyakit dengan jantung mengalami gangguan seperti

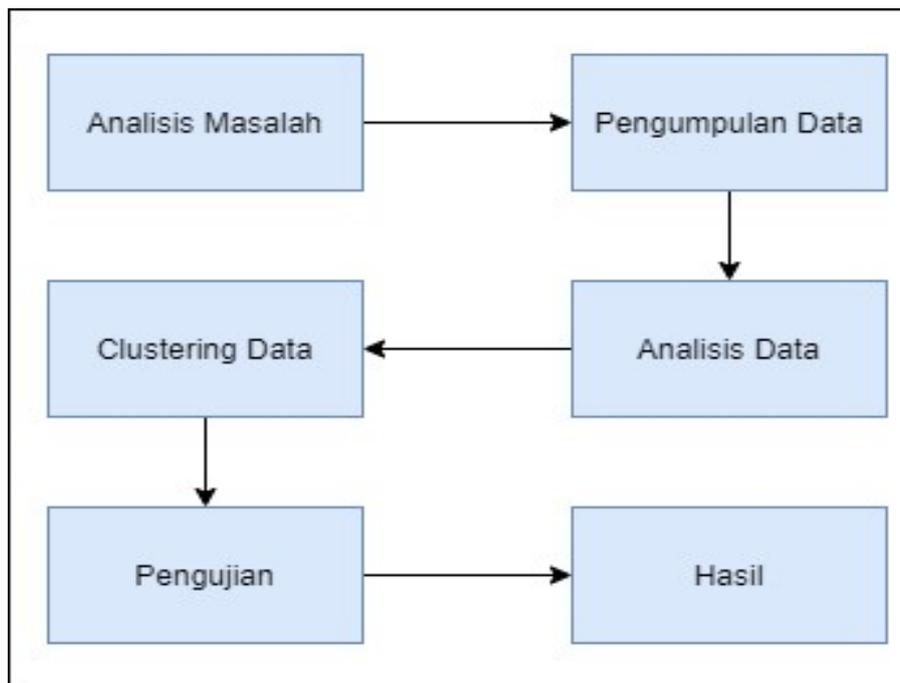
aritmia, gagal jantung, jantung koroner, dan sebagainya. Aritmia merupakan penyakit karena terjadi gangguan terhadap irama jantung sehingga menyebabkan irama atau denyut jantung tidak normal. Gagal jantung merupakan penyakit karena kegagalan otot jantung dalam memompa darah secara menyeluruh ke seluruh tubuh. Penyakit jantung koroner merupakan penyakit yang disebabkan penyempitan pembuluh darah di jantung. Pada tahun 2012, data dari WHO bahwa 17,5 juta orang di dunia meninggal dunia dikarenakan penyakit kardiovaskular. Penyakit jantung merupakan salah satu jenis dari penyakit kardiovaskular yang ikut menyumbang 31% dari total 56,5 juta kasus kematian di dunia (Jabbar MA, dkk, 2013).

Penyakit jantung terjadi ketika darah yang mengalir ke otot jantung berhenti sehingga menyebabkan gangguan jantung. Hal ini menyebabkan adanya kebutuhan mendefinisikan sistem pendukung keputusan yang membantu dokter dalam mengambil keputusan untuk mengambil tindakan pencegahan terhadap penderita penyakit jantung. Berdasarkan kebutuhan tersebut, maka perlu didesain sistem untuk melakukan klasifikasi dan diagnosis tentang kesehatan jantung. Metode yang dapat digunakan adalah menggunakan *artificial intelligence* (AI). Metode AI didasarkan pada *neural network* (NN) yang jenisnya beragam dan disesuaikan dengan kebutuhan (Sari AP, dkk, 2022). Dalam makalah ini, metode NN yang digunakan adalah algoritma K-NN (K- Nearest Neighbor) yang tujuannya dapat menyelesaikan sistem klasifikasi (pengelompokan) dan diagnosis kesehatan jantung.

K-NN merupakan metode klasifikasi yang menggunakan hubungan kedekatan antara suatu jarak data dengan data lain (Hasran, 2020). Metode K-NN menggunakan data berdimensi dengan menghitung jarak dari suatu data ke data lain. Nilai jarak digunakan ialah nilai kedekatan antara data yang dilatih dan data yang diuji. Nilai K pada metode K-NN merupakan K-data terdekat dari data yang diuji (Mei L, 2014). Dalam mempermudah sistem klasifikasi dan diagnosis, maka dibuat sistem pengelompokan kesehatan Jantung.

2. Metodologi

Metodologi pada penelitian ini terdiri atas beberapa tahapan yang dilakukan secara berurutan yang dapat ditunjukkan dalam Gambar 1. Tahapan metodologi penelitian, antara lain: analisis masalah, pengumpulan data, analisis data, clustering data, pengujian, hasil, dan ditarik kesimpulan. Setiap tahapan metodologi dijelaskan pada tiap subbab selanjutnya.



Gambar 1 Tahapan Penelitian

2.1 Analisis Masalah

Analisis masalah ialah suatu tahapan dalam proses menganalisis yang berguna untuk mengetahui sebab dari suatu masalah yang ada di penelitian.

2.2 Pengumpulan Data

Pengumpulan data ialah suatu tahapan yang berguna untuk memperoleh informasi berdasarkan dari sumber yang *valid*, contoh: buku, *repository*, jurnal ilmiah, *dataset*, *website*, dan lain lain. Hal ini bermanfaat sebagai penunjang dalam sebuah penelitian (Sihananto AN, dkk, 2022).

2.3 Analisis Data

Analisis data ialah suatu tahapan dalam mencermati kumpulan data. Kemudian proses selanjutnya mencari data yang cocok dipakai dalam suatu penelitian, contoh: melaksanakan observasi pada data tertentu.

2.4 Clustering Data

Clustering data ialah suatu tahapan yang mengelompokkan data menjadi beberapa bagian dengan tujuan mendapatkan hasil dengan tipe sama (Lestari EP dkk, 2020)(Rheza MA, dkk, 2020). Tahap ini menggunakan metode K-NN *Clustering* dengan menemukan klasterisasi K pada data latih yang terdekat dengan data uji. Selain itu, metode ini juga melaksanakan pemeriksaan label pada kelas tertentu yang masih pada ruang lingkup data tertentu. Pada metode K-NN terdapat beberapa perhitungan dan yang paling umum dipakai ialah perhitungan didasarkan pada jarak terdekat (Noviansyah MR, dkk, 2018). Fungsi Penghitungan dari metode

K-NN, adalah sebagai berikut:

$$d_{Euclidean}(x, y) = \sqrt{\sum_1 (x_i - y_i)^2} \quad (1)$$

2.5 Pengujian

Pengujian ialah suatu tahapan untuk menetapkan tingkat validitas dan keakuratan dari suatu sistem sudah berfungsi secara baik atau tidak. Tahap ini memakai metode bawaan dari *library* K-NN yaitu fungsi *predict*.

3. Hasil dan Pembahasan

3.1 Analisis Masalah

Permasalahan yang dibahas pada makalah ini ialah kasus kesehatan jantung yang sampai detik ini masih kedapatan di belahan bumi manapun. Dalam halantisipasi beberapa kejadian diluar dugaan, maka memerlukan pemrosesan *clustering*. Hal ini dilaksanakan untuk memahami bagaimana kondisi jantung seseorang terhadap berbagai faktor kesehatan.

3.2 Pengumpulan Data

Dataset (sumber data utama) yang dipakai dalam makalah ini adalah data kesehatan jantung yang sumbernya didapatkan dari *platform* kaggle.com. *Dataset* yang diambil untuk dilakukan analisis pada penelitian ini adalah 13 atribut kategori dari 76 atribut kategori yang terdapat pada kaggle.com dengan masing-masing data terdiri atas 303 data.

3.3 Analisis Data

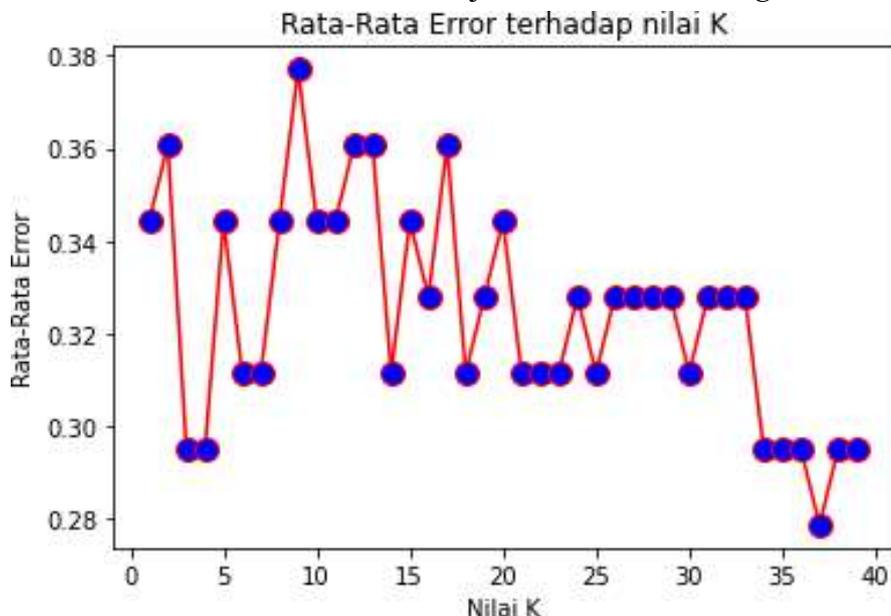
Data sampel yang diproses merupakan hasil dari data responden yang sudah dikelompokkan sehingga menjadi dataset dan disediakan pada kaggle.com. Berdasarkan dataset yang disediakan kaggle.com diambil 13 atribut kategori data saja yang disesuaikan dengan kebutuhan yang ditunjukkan pada Tabel 1.

Tabel 1 Datasheet

NO	Atribut Kategori Data	Jumlah Data
1.	Age	303
2.	Sex	303
3.	Chest Pain	303
4.	Resting Blood Pressure	303
5.	Cholesterol	303
6.	Fasting Blood Sugar	303
7.	Resting Electrocardiographic Result	303
8.	Maximum Heart Rate	303
9.	Exercise Induced Angina	303
10.	Oldpeak	303
11.	Slope ST Segment	303
12.	Number of Major Vessels	303
13.	Heart Conditions	303

3.4 K-NN Clustering

Penentuan jumlah *cluster* (k) optimal berdasarkan pada data dilaksanakan klusterisasi dengan metode K-NN yang memakai fungsi bawaan *library* K-NN. Kemudian dilakukan *Clustering* menggunakan metode K-NN. Gambar 3 menunjukkan hasil *Clustering* K-NN.



Gambar 2 Hasil Clustering K-NN

3.5 Pengujian Data

Pengujian data perlu dilakukan untuk menghitung tingkat keakuratan *clustering*. Penghitungan tingkat keakuratan *clustering* menggunakan fungsi *predict*

pada *library* K-NN yang ditunjukkan pada Gambar 5. Pengujian menggunakan metode K-NN diperoleh hasil akurasi yaitu: 0.70 atau 70 %.

```
y_pred = classifier.predict(X_test)
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.70	0.94	0.81	33
2	0.00	0.00	0.00	4
3	0.71	0.50	0.59	24
accuracy			0.70	61
macro avg	0.47	0.48	0.46	61
weighted avg	0.66	0.70	0.67	61

Gambar 3 Pengujian Fungsi Predict

4. Kesimpulan dan Saran

Berdasarkan dari hasil penelitian, maka dapat ditarik suatu kesimpulan bahwasanya permasalahan yang diselesaikan yaitu: pengelompokan kasus kesehatan jantung berdasarkan 13 atribut kategori data dan menggunakan metode *K-Nearest Neighbours clustering*. Hal tersebut didapatkan 3 *cluster* yaitu *cluster 0 (C0)*, *cluster 2 (C1)*, dan *cluster 3 (C2)*. Setelah melewati pengujian atau evaluasi didapatkan tingkat keakuratan metode K-NN senilai 0.70 atau 70 %. Saran bagi penelitian selanjutnya, ialah: menggunakan kombinasi dua metode, misal K-NN dan *K-Means Clustering* sehingga nilai keakuratan meningkat.

Daftar Pustaka

Jabbar MA, Deekshatulua BL, Chandra P. (2013). Classification of heart disease using K-nearest neighbor and genetic algorithm. *Procedia Technology* 10 : 85 – 94.

Sari AP, Suzuki H, Kitajima T, Yasuno T, Prasetya DA, Arifuddin R. (2022). Short-Term Wind Speed and Direction Forecasting by 3DCNN and Deep Convolutional LSTM. *IEEE Transactions on Electrical and Electronic Engineering* 2022; 17: 1620–1628.

Hasran. (2020). Klasifikasi penyakit jantung menggunakan metode k-nearest neighbor. *Indonesian Journal of Data and Science* Vol 1, pp. 06-10.

Mei L. (2014). Penerapan algoritma klasifikasi nearest neighbor (k-nn) untuk mendeteksi penyakit jantung. *Faktor Exacta* 7(4): 366-371.

Sihananto AN, Sari AP, Khariono H, Fernanda RA, Wijaya DCM. (2022). Implementasi metode k-means untuk pengelompokan kasus covid 19 tingkat provinsi di indonesia. *Jurnal Informatika dan Sistem Informasi*, 3(1), 76-85.

Lestari E P, Mahmudi A, Achmadi S. (2020). Penerapan metode k-means untuk proses penentuan golongan penerima zakat (mustahiq) di yayasan dana sosial al-falah malang. *JATI (Jurnal Mhs. Tek. Inform., vol. 4, no. 1, pp. 389–396.*

Rheza MA, Metandi F. (2020). Implementasi metode k-means clustering untuk penentuan jenis komentar pada tweet pssi. *Just TI (Jurnal Sains Terap. Teknol. Informasi)*, vol. 12, no. 2, p. 73, 2020.

Noviansyah MR, Rismawan T, Midyanti DM. (2018). Penerapan data mining menggunakan metode k-nearest neighbor untuk klasifikasi indeks cuaca kebakaran berdasarkan data AWS (automatic weather station) (studi kasus: kabupaten kubu raya). *Jurnal Coding*, vol. 06, No. 2, hal 48-56, 2018.