

Penerapan Metode *K-Nearest Neighbor* untuk Mengklasifikasikan Penyakit *Cardiovascular*

Nina Kurnia Rahayu *, Nur Azizah Komara Rifai

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

ninakrahayu@gmail.com, nur.azizah@unsiba.ac.id

Abstract. Cardiovascular is one of the organs of the human body that plays a role in the circulatory system. Cardiovascular disease is a condition in which the heart is unable to function properly. It is the leading cause of death worldwide each year. At least three-quarters of the world's deaths from cardiovascular disease occur in low- and middle-income countries. These communities often do not benefit from health care programs. Most cardiovascular diseases can be prevented by addressing behavioral risk factors such as smoking, unhealthy diets and obesity, physical inactivity and harmful use of alcohol. Public health policies that create an environment that supports healthy, affordable, and accessible food choices should encourage people to adopt and maintain healthy behaviors. Therefore, this study aims to classify which factors cause cardiovascular disease by applying the K-Nearest Neighbor (K-NN) method. K-NN is one of the machine learning algorithms that is relatively easy to implement and produce estimates that can be used for better decision making in the future. The data used in this study consists of 12 variables with a total of 70000 data for each variable. By dividing the data with a proportion of 70% for training data and 30% for test data, the best K is obtained, namely K = 5. The model evaluation shows an accuracy value of 60.03%, which illustrates the success rate of the model in predicting cardiovascular disease categories based on observed parameters. These results illustrate that the K-Nearest Neighbor method is able to provide predictions with fairly good accuracy in identifying cardiovascular disease.

Keywords: K-NN, Euclidean Distance, Cardiovascular.

Abstrak. *Cardiovascular* merupakan salah satu organ tubuh manusia yang berperan dalam sistem peredaran darah. Penyakit *cardiovascular* adalah suatu kondisi di mana jantung tidak mampu menjalankan fungsinya dengan baik. Secara global, penyakit ini merupakan penyebab kematian paling umum setiap tahunnya. Setidaknya tiga perempat kematian akibat penyakit *cardiovascular* di seluruh dunia terjadi di negara-negara berpenghasilan rendah dan menengah. Masyarakat ini seringkali tidak mendapatkan manfaat dari program layanan kesehatan. Sebagian besar penyakit *cardiovascular* dapat dicegah dengan mengatasi faktor risiko perilaku seperti merokok, pola makan tidak sehat dan obesitas, kurang aktivitas fisik, dan konsumsi alkohol yang berbahaya. Kebijakan kesehatan yang menciptakan lingkungan yang mendukung pilihan makanan yang sehat, terjangkau dan mudah diakses harus mendorong masyarakat untuk mengadopsi dan mempertahankan perilaku sehat. Oleh karena itu, penelitian ini bertujuan untuk mengklasifikasi faktor apa saja yang menjadi penyebab penyakit *cardiovascular* dengan menerapkan metode *K-Nearest Neighbor* (K-NN). K-NN adalah salah satu algoritma *machine learning* yang relatif mudah dan dapat diimplementasikan serta dapat menghasilkan estimasi yang dapat digunakan untuk pengambilan keputusan yang lebih baik di masa mendatang. Data yang digunakan pada penelitian ini terdiri dari 12 variabel dengan jumlah data sebanyak 70000 pada setiap variabel. Melalui pembagian data dengan proporsi 70% untuk data *train* dan 30% untuk data *test*, diperoleh K terbaik yaitu K = 5. Evaluasi model menunjukkan nilai akurasi sebesar 60,03% yang menggambarkan tingkat keberhasilan model dalam memprediksi kategori penyakit *cardiovascular* berdasarkan parameter-parameter yang diamati. Hasil ini memberikan gambaran bahwa metode K-NN mampu memberikan prediksi dengan akurasi yang cukup baik dalam mengidentifikasi penyakit *cardiovascular*.

Kata Kunci: K-NN, Euclidean Distance, Cardiovascular.

A. Pendahuluan

Cardiovascular merupakan salah satu organ tubuh manusia yang berperan dalam sistem peredaran darah. Penyakit cardiovascular adalah suatu kondisi di mana jantung tidak mampu menjalankan fungsinya dengan baik. Secara global, penyakit cardiovascular merupakan penyebab kematian paling umum setiap tahunnya. Diperkirakan 17,9 juta orang meninggal karena penyakit cardiovascular pada tahun 2019, yang merupakan 32% dari seluruh kematian global. Dari 85% kematian ini disebabkan oleh serangan jantung dan stroke. Pada tahun 2019 ada 38% dari 17 juta kematian dini (di bawah usia 70 tahun) akibat penyakit tidak disebabkan oleh penyakit cardiovascular.

Setidaknya tiga perempat kematian akibat penyakit cardiovascular di seluruh dunia terjadi di negara-negara berpenghasilan rendah dan menengah. Masyarakat ini seringkali tidak mendapatkan manfaat dari program layanan kesehatan primer yang ditujukan untuk mendeteksi dini dan menangani orang-orang dengan faktor risiko cardiovascular (Salma & Yanti, 2024). Masyarakat di negara-negara berpenghasilan rendah dan menengah yang menderita penyakit cardiovascular dan penyakit tidak menular lainnya mempunyai akses yang lebih kecil terhadap layanan kesehatan yang efektif dan adil untuk memenuhi kebutuhan mereka. Akibatnya, penyakit ini sering kali terlambat terdeteksi bagi banyak orang di negara-negara tersebut, dan banyak orang meninggal karena penyakit cardiovascular dan penyakit tidak menular lainnya pada usia paling produktif dan pada usia yang lebih muda.

Sebagian besar penyakit cardiovascular dapat dicegah dengan mengatasi faktor risiko perilaku seperti merokok, pola makan tidak sehat dan obesitas, kurang aktivitas fisik, dan konsumsi alkohol yang berbahaya. Dampak dari faktor risiko perilaku dapat dilihat pada individu seperti peningkatan tekanan darah, peningkatan gula darah, peningkatan kadar lipid darah, dan kelebihan berat badan atau obesitas. “Faktor risiko menengah” ini dapat diukur di fasilitas layanan kesehatan primer dan menunjukkan peningkatan risiko serangan jantung, stroke, gagal jantung, dan komplikasi lainnya (Islami, 2008).

Berhenti merokok, mengurangi garam dalam makanan, makan lebih banyak buah dan sayuran, melakukan aktivitas fisik secara teratur dan menghindari konsumsi alkohol yang berbahaya dapat mengurangi risiko penyakit cardiovascular. Kebijakan kesehatan yang menciptakan lingkungan yang mendukung pilihan makanan yang sehat, terjangkau dan mudah diakses harus mendorong masyarakat untuk mengadopsi dan mempertahankan perilaku sehat. Selain itu, pengobatan hipertensi, diabetes, dan kadar lipid darah tinggi penting untuk mengurangi risiko cardiovascular dan mencegah serangan jantung dan stroke pada penderita penyakit ini.

Penggunaan metode K-Nearest Neighbor (K-NN) untuk mengklasifikasi faktor apa saja yang menjadi penyebab penyakit cardiovascular memiliki kemampuan dalam melakukan prediksi berbasis pola data historis. K-NN adalah salah satu algoritma machine learning yang relatif mudah dan dapat di implementasikan, mampu menggunakan data historis (Esesiawati, 2017) tentang faktor yang mempengaruhi penyakit cardiovascular dengan memanfaatkan data sekunder seperti data pasien yang mempunyai penyakit cardiovascular, K-NN dapat menghasilkan estimasi yang dapat digunakan untuk pengambilan keputusan yang lebih baik di masa yang akan datang.

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut: “Bagaimana penerapan metode K-Nearest Neighbor untuk mengklasifikasikan penyakit cardiovascular?”. Selanjutnya, tujuan dalam penelitian ini yaitu untuk mengetahui penerapan metode K-Nearest Neighbor untuk mengklasifikasikan penyakit cardiovascular.

B. Metode

Peneliti menggunakan metode *K-Nearest Neighbor* yang merupakan metode nonparametrik untuk klasifikasi berdasarkan tetangga terdekat. Algoritma *K-Nearest Neighbors* digunakan untuk menentukan jarak antara data yang dievaluasi (data *training*) dan himpunan k tetangga terdekat pada data baru (data *testing*). Pengukuran jarak antara data *training* dan data *testing* yang paling sering digunakan adalah *euclidean distance*. *Euclidean distance* di definisikan pada persamaan 1 (Sreemathy & Balamruga, 2012) :

$$Dist(X, Y) = \sqrt{\sum_{i=1}^D (X_i - Y_i)^2} \quad \dots (1)$$

di mana:

$Dist(X, Y)$	: jarak antara objek
X_i	: sampel data
Y_i	: data uji
D	: dimensi data
i	: variabel data

Jenis penelitian yang di gunakan pada penelitian ini yaitu deskriptif kuantitatif dengan data yang di gunakan merupakan data sekunder. Data sekunder yaitu data yang tidak dikumpulkan secara langsung, melainkan diperoleh dari data yang sudah tersedia. Data yang di gunakan pada penelitian ini yaitu data penyakit *cardiovascular* dengan variabel data sebanyak 12 variabel dan jumlah data sebanyak 70000 data pada setiap variabel yang di peroleh dari website *Kaggle*. Adapun teknik pengumpulan data pada penelitian ini yaitu teknik dokumenter. Teknik dokumenter adalah teknik pengumpulan data menggunakan sumber dokumen tertulis yang sudah tersedia (Rachma, 2022). Penelitian ini menggunakan bantuan *software Jupyter Notebook* dan *tools Microsoft Excel*. Pada penelitian ini, data dibagi menjadi rasio sebanyak 30% pada data *testing* dan 70% pada data *training*. Berikut instrumen penelitian yang di gunakan pada penelitian ini:

1. Variabel dependent (Y) yang merupakan label/kelas pada peneltian ini berdasarkan ada atau tidaknya penyakit *cardiovascular* dengan label/kelas kelas 0 (tidak ada penyakit *cardiovascular*) dan kelas 1 (ada penyakit *cardiovascular*).
2. Variabel indpendent (X) yang terdiri dari usia (X_1) dalam hari, tinggi badan (X_2) dalam cm, berat badan (X_3) dalam kg, jenis kelamin (X_4) (1 = perempuan, 2 = laki-laki), tekanan darah sistolik (X_5), tekanan darah diastolik (X_6), kolesterol (X_7) (1 = normal, 2 = di atas normal, 3 = jauh di atas normal), tekanan gula darah (X_8) (1 = normal, 2 = di atas normal, 3 = jauh di atas normal), merokok (X_9) (0 = tidak, 1= ya), asupan alkohol (X_{10}) (0 = tidak, 1 = ya), aktivitas fisik (X_{11}) (0 = tidak, 1 = ya).

Menurut Asha Gowda (2012) “K-NN sangat mudah di implementasikan dan mudah untuk membenarkan hasil KNN. Namun, KNN juga memiliki beberapa kekurangan seperti biaya komputasi yang tinggi karena harus menghitung jarak setiap contoh uji ke semua sampel pelatihan, memerlukan memori besar yang sebanding dengan ukuran set pelatihan, tingkat akurasi rendah pada set data multidimensi dengan fitur yang tidak relevan, tidak ada aturan praktis untuk menentukan nilai parameter K ”.

Algoritma *K-Nearest Neighbor* dituliskan sebagai berikut (Asha Gowda, 2012): (1) Menentukan nilai k tetangga terdekat, nilai k harus memenuhi syarat $k < N$ yang mana N adalah jumlah kumpulan data. (2) Menhgitung jarak antara data baru dan data latih, pengukuran menggunakan jarak *euclidean*. (3) Mengurutkan jarak dari yang terdekat. (4) Mengecek kelas k tetangga terdekat. (5) Mendapatkan kelas data baru yang biasanya merupakan kelas tetangga terdekat.

C. Hasil Penelitian dan Pembahasan

Analisis Deskriptif

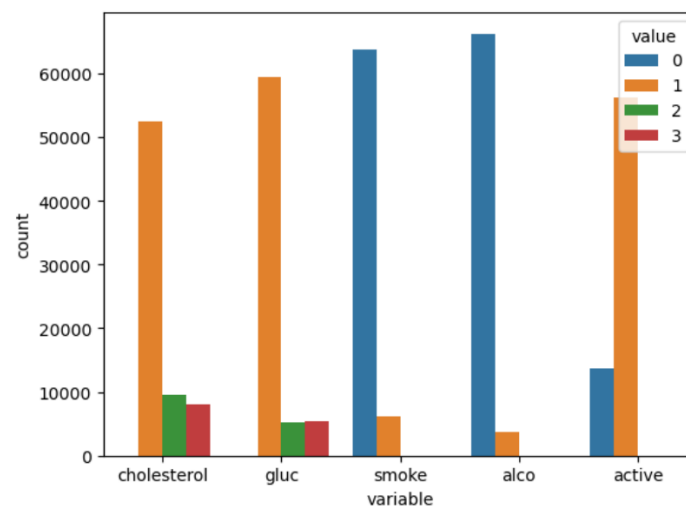
Tahap analisis deksriptif pada penelitian ini bertujuan untuk mendapatkan gambaran yang jelas tentang karakteristik data penyakit *cardiovascular*. Variabel dependent pada penelitian ini adalah ada atau tidaknya penyakit *cardiovascular*. Sedangkan variabel independent terdiri dari usia (hari), tinggi badan, berat badan, jenis kelamin, tekanan darah sistolik, tekanan darah diastolik, kolestrol, tekanan gula darah, merokok, asupan alkohol dan aktivitas fisik.

Tabel 1. Analisis Deskriptif

Variabel	N	Mean	Std. Deviasi	Minimum	Maximum
Y	70000	0,4997	0,5000	0	1
X_1	70000	19468,8658	2467,2517	10798	23713
X_2	70000	164,3592	8,2101	55	250
X_3	70000	74,2057	14,3958	10	200
X_4	70000	1,3496	0,4768	1	2
X_5	70000	128,8173	154,0114	-150	16020
X_6	70000	96,6304	188,4725	-70	11000
X_7	70000	1,3669	0,6803	1	3

Variabel	N	Mean	Std. Deviasi	Minimum	Maximum
X_8	70000	1,2265	0,5723	1	3
X_9	70000	0,0881	0,2835	0	1
X_{10}	70000	0,0538	0,2256	0	1
X_{11}	70000	0,8037	0,3972	0	1

Berdasarkan Tabel 1. dapat dilihat bahwa variabel usia (X_1) mempunyai nilai rata-rata sebesar 19468,8658 hari, karena nilai standar deviasinya 2467,2517 lebih kecil dari nilai rata-rata artinya variabel usia (X_1) bersifat homogen, serta nilai minimum sebesar 10798 dan maksimum sebesar 23713. Selanjutnya pada variabel tinggi badan (X_2) mempunyai nilai rata-rata sebesar 164,3592 karena nilai standar deviasinya 8,2101 lebih kecil dari nilai rata-rata artinya variabel tinggi badan (X_2) bersifat homogen, serta nilai minimum sebesar 55 dan maksimum sebesar 250. Variabel berat badan (X_3) mempunyai nilai rata-rata sebesar 74,2057 karena nilai standar deviasinya 14,3958 lebih kecil dari nilai rata-rata artinya variabel berat badan (X_3) bersifat homogen, serta nilai minimum sebesar 10 dan maksimum sebesar 200. Variabel Variabel tekanan darah sistolik (X_5) mempunyai nilai rata-rata sebesar 128,8173 karena nilai standar deviasinya 154,0114 lebih kecil dari nilai rata-rata artinya variabel tekanan darah sistolik (X_5) bersifat homogen, serta nilai minimum sebesar -150 dan maksimum sebesar 16020. Variabel tekanan darah diastolik (X_6) mempunyai nilai rata-rata sebesar 96,6304 karena nilai standar deviasinya 188,4725 lebih kecil dari nilai rata-rata artinya variabel tekanan darah diastolik (X_6) bersifat homogen, serta nilai minimum sebesar -70 dan maksimum sebesar 11000.



Gambar 1. Analisis Deskriptif

Gambar 1. menunjukkan bahwa pada variabel kolestrol (X_7) kategori normal merupakan kategori kolestrol dengan frekuensi paling banyak, artinya penderita penyakit *cardiovascular* memiliki tingkat kolestrol yang normal. Variabel tekanan gula darah (X_8) kategori normal merupakan kategori dengan frekuensi paling banyak, artinya penderita penyakit *cardiovascular* memiliki tingkat gula darah yang normal. Variabel merokok (X_9) kategori tidak merokok merupakan kategori dengan frekuensi paling banyak, artinya penderita penyakit *cardiovascular* rata-rata bukan perokok. Variabel asupan alkohol (X_{10}) kategori tidak minum alkohol merupakan kategori dengan frekuensi paling banyak, artinya penderita penyakit *cardiovascular* rata-rata bukan peminum alkohol. Sedangkan, variabel aktivitas fisik (X_{11}) kategori beraktivitas merupakan kategori dengan frekuensi paling banyak, artinya penderita penyakit *cardiovascular* rata-rata beraktivitas fisik.

Standarisasi Data

Tahap standarisasi data dilakukan karena variabel-variabel pada data tidak memiliki rentang nilai yang sama. Hasil perhitungan standarisasi data sebagai berikut:

Hasil stadarisasi untuk variabel X_1

$$X_{11} = \frac{X_1 - X_{\min}}{X_{\max} - X_{\min}} = \frac{18393 - 10798}{23713 - 10798} = 0,58807$$

$$X_{12} = \frac{X_2 - X_{\min}}{X_{\max} - X_{\min}} = \frac{20228 - 10798}{23713 - 10798} = 0,73015$$

$$\vdots$$

$$X_{170000} = \frac{X_{70000} - X_{\min}}{X_{\max} - X_{\min}} = \frac{20540 - 10798}{23713 - 10798} = 0,75431$$

Hasil standarisasi untuk variabel X_2

$$X_{21} = \frac{X_1 - X_{\min}}{X_{\max} - X_{\min}} = \frac{2 - 55}{250 - 55} = -0,27179$$

$$X_{22} = \frac{X_2 - X_{\min}}{X_{\max} - X_{\min}} = \frac{1 - 55}{250 - 55} = -0,27692$$

$$\vdots$$

$$X_{270000} = \frac{X_{70000} - X_{\min}}{X_{\max} - X_{\min}} = \frac{1 - 55}{250 - 55} = -0,27692$$

Hasil standarisasi untuk variabel X_3

$$X_{31} = \frac{X_1 - X_{\min}}{X_{\max} - X_{\min}} = \frac{168 - 10}{200 - 10} = 0,83157$$

$$X_{32} = \frac{X_2 - X_{\min}}{X_{\max} - X_{\min}} = \frac{156 - 10}{200 - 10} = 0,76842$$

$$\vdots$$

$$X_{370000} = \frac{X_{70000} - X_{\min}}{X_{\max} - X_{\min}} = \frac{170 - 10}{200 - 10} = 0,84210$$

Hasil standarisasi untuk variabel X_4

$$X_{41} = \frac{X_1 - X_{\min}}{X_{\max} - X_{\min}} = \frac{62 - 1}{2 - 1} = 61$$

$$X_{42} = \frac{X_2 - X_{\min}}{X_{\max} - X_{\min}} = \frac{85 - 1}{2 - 1} = 84$$

$$\vdots$$

$$X_{470000} = \frac{X_{70000} - X_{\min}}{X_{\max} - X_{\min}} = \frac{72 - 1}{2 - 1} = 71$$

Hasil standarisasi untuk variabel X_5

$$X_{51} = \frac{X_1 - X_{\min}}{X_{\max} - X_{\min}} = \frac{110 - (-150)}{16020 - (-150)} = 0,01608$$

$$X_{52} = \frac{X_2 - X_{\min}}{X_{\max} - X_{\min}} = \frac{156 - (-150)}{200 - (-150)} = 0,01793$$

$$\vdots$$

$$X_{570000} = \frac{X_{70000} - X_{\min}}{X_{\max} - X_{\min}} = \frac{120 - (-150)}{16020 - (-150)} = 0,01669$$

Dan seterusnya sampai variabel X_{12} .

Tabel 2. Data Hasil Standarisasi

id	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
0	0,58807	-0,27179	0,83157	61	0,01607	0,01355	0	0	0	0	1	0
1	0,73015	-0,27692	0,76842	84	0,01793	0,01445	1	0	0	0	1	1
2	0,62400	-0,27692	0,81578	63	0,01731	0,01264	1	0	0	0	0	1
3	0,52845	-0,27179	0,83684	81	0,01855	0,01535	0	0	0	0	1	1
4	0,51691	-0,27692	0,76842	55	0,01546	0,01174	0	0	0	0	0	0
8	0,86070	-0,27692	0,74210	66	0,01669	0,01355	0,5	0,5	0	0	0	0
9	0,87611	-0,27692	0,77368	92	0,01731	0,01355	1	0	0	0	1	0

id	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
70000	0,51691	-0,27692	0,76842	55	0,01546	0,01174	0	0	0	0	0	0

Tahap Preprocessing

Pada penelitian ini tahap preprocessing dilakukan untuk menghilangkan variabel 'id' pada data, karena variabel tersebut tidak akan digunakan dalam tahap pengujian.

Implementasi Algoritma *K-Nearest Neighbor*

Pengklasifikasian dengan algoritma *K-Nearest Neighbor* (K-NN) data dibagi menjadi dua proses, yaitu proses *training* dan *testing*. Pada penelitian ini, data dibagi menjadi rasio sebanyak 30% pada data *testing* dan 70% pada data *training*. Perhitungan manual dengan menggunakan bantuan *Microsoft Excel* sebagai berikut:

Tabel 3. Hasil Implementasi Algoritma *K-Nearest Neighbor*

id	d	rank
0	5,319739	48553
1	18,06407	15494
2	3,507419	56544
3	15,10996	19592
4	11,19476	27345
8	1,688673	69377
9	26,04523	7181
⋮	⋮	⋮
99999	5,249932	48939

Selanjutnya akan dilakukan penentuan label, satu contoh data yang belum diketahui labelnya. Berikut merupakan langkah-langkah perhitungan manual algoritma *K-Nearest Neighbor* (K-NN):

1. Menetapkan nilai K sebagai parameter banyaknya jumlah tetangga terdekat yang akan digunakan.
2. Menghitung jarak antara objek yang belum diketahui labelnya terhadap semua objek yang ada di data latih.
3. Mengurutkan hasil jarak dari yang paling kecil ke besar dan menentukan jarak terdekat berdasarkan nilai K yang sudah ditentukan sebelumnya.
4. Setelah diurutkan, jarak *euclidean* terkecil berada pada id ke-618 sebesar 133,0163 yang termasuk kedalam kategori bukan penderita penyakit *cardiovascular*. Maka objek baru yang di uji tersebut bukan merupakan penderita penyakit *cardiovascular*.

Evaluasi Model

Evaluasi model dilakukan dengan menggunakan *confusion matrix* untuk mengukur kinerja model yang telah di buat. Berikut hasil dari evaluasi model:

Tabel 4. Hasil Evaluasi Model

Confusion Matrix		Nilai Aktual	
Nilai Prediksi	Tidak ada	6438	4101
	Ada	4293	6168

Tabel 4. menunjukkan bahwa dapat mengklasifikasikan 6438 bukan penderita penyakit *cardiovascular* yang terprediksi benar sebagai bukan penderita penyakit tersebut dan 6168 penderita penyakit *cardiovascular* yang terprediksi benar sebagai penderita penyakit *cardiovascular*. Kemudian terdapat 4293 bukan penderita penyakit *cardiovascular* yang terprediksi salah kedalam penderita penyakit *cardiovascular* dan 4101 penderita penyakit *cardiovascular* yang terprediksi salah kedalam

bukan penderita penyakit *cardiovascular*. Adapun nilai akurasi diperoleh melalui perhitungan sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Akurasi = \frac{6438 + 6168}{6438 + 6168 + 4293 + 4101} \times 100\%$$

$$Akurasi = 60,03\%$$

Berdasarkan hasil klasifikasi K-Nearest Neighbor dengan menggunakan nilai K=5 diperoleh nilai akurasi sebesar 60,03%

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan hasil penelitian bahwa pada penelitian ini dilakukan analisis dengan menggunakan algoritma *K-Nearest Neighbor* dengan data yang diperoleh dari website *Kaggle* yaitu data penyakit *cardiovascular*. Data ini memiliki 12 variabel dengan satu variabel sebagai variabel independent dan 11 variabel sebagai variabel dependent. Melalui pembagian data dengan proporsi 70% untuk data *train* dan 30% untuk data *test*, algoritma *K-Nearest Neighbor* diterapkan dalam proses klasifikasi, dan diperoleh K terbaik yaitu K = 5. Evaluasi model menunjukkan nilai akurasi sebesar 60,03% yang menggambarkan tingkat keberhasilan model dalam memprediksi kategori penyakit *cardiovascular* berdasarkan parameter-parameter yang diamati. Hasil ini memberikan gambaran bahwa metode *K-Nearest Neighbor* mampu memberikan prediksi dengan akurasi yang cukup baik dalam mengidentifikasi penyakit *cardiovascular*.

Ucapan Terimakasih

Terima kasih kepada Allah SWT yang telah memberikan rahmatnya sehingga peneliti dapat menyelesaikan penelitian ini dengan lancar. Selanjutnya terima kasih kepada semua pihak yang telah membantu dalam menyelesaikan dan memberikan saran untuk penelitian ini serta kepada pembaca yang sudah bersedia meluangkan waktunya untuk membaca artikel penelitian ini. Diharapkan artikel penelitian ini dapat memberikan informasi tambahan juga manfaat bagi berbagai pihak yang membacanya.

Daftar Pustaka

- Salma, G., & Yanti, T. S. (2024). Pemodelan Intervensi untuk Meramalkan Jumlah Penumpang Pesawat Domestik Bandara Internasional Sultan Hasanuddin Makassar. *Jurnal Riset Statistika*, 67–74. <https://doi.org/10.29313/jrs.v4i1.3892>
- Asha Gowda Karegowda, M.A. Jayaram, & A.S. Manjunath. (2012). Cascading kmeans clustering and k-nearest neighbor classifier for categorization of diabetic patients. *International Journal of Engineering and Advanced Technology*, 1(3), 147–151. <http://www.galaxy.gmu.edu/interface/I01/I2001Proceedings/Jbreault>.
- Esesiawati, O. (2017). Peramalan Arus Lalu Lintas Jangka Pendek Menggunakan Metode K-Nearest Neighbor (Studi Kasus : Jalan Basuki Rahmat Surabaya).
- Islami, A. U. (2008). Faktor-faktor risiko penyakit kardiovaskular, 13-33.
- Lall, U., & Sharma, A. (1996). A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Research*, 32(3), 679–693. <https://doi.org/10.1029/95WR02966>.
- Meade, N. (2002). A comparison of the accuracy of short term foreign exchange forecasting methods. *International Journal of Forecasting*, 18(1), 67–83. [https://doi.org/10.1016/S0169-2070\(01\)00111-X](https://doi.org/10.1016/S0169-2070(01)00111-X).

- Putri, I. P. (2021). Analisis Performa Metode K-Nearest Neighbor (KNN) dan Crossvalidation pada Data Penyakit Cardiovascular. *Indonesian Journal of Data and Science*, 2(1), 21- 28.
- Rachma, C. A. (2022). *Implementasi Algoritma K-Nearest Neighbor dalam penentuan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi Jawa Timur* (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).
- Sreemathy, J., & Balamurugan, P. S. (2012). An efficient text classification using knn and naive bayesian. *International Journal on Computer Science and Engineering*, 4(3), 392.
- Wang, H., Xu, P., & Zhao, J. (2021). Improved KNN algorithms of spherical regions based on clustering and region division. *Alexandria Engineering Journal*, 1–15. <https://doi.org/10.1016/j.aej.2021.09.004>.
- WHO, 2021. Cardiovascular Diseases (CVDs) (Online), (<https://www.who.int/news-room/fact-sheets/detail/cardiovascular-dis>).