

Penerapan Algoritma *K-Nearest Neighbour* pada Penyakit Jantung

Rohmatul Maula¹, Saiful Bahri².

^{1,2}Program Studi Matematika, Fakultas Sains dan Matematika, Universitas Islam Negeri Sunan Ampel Surabaya, Jl. Dr. Ir. H. Soekarno No. 682, Gn. Anyar, Surabaya, Indonesia

Korespondensi; Rohmatul Maula, Email: rohmadan06@gmail.com

Abstrak

Jantung adalah organ paling penting yang dimiliki oleh manusia. Apabila terjadi kerusakan yang disebabkan penyempitan pembuluh darah, maka akan menyebabkan penyakit jantung. Penyakit jantung menempati urutan pertama kematian utama di Indonesia. Penelitian ini bertujuan untuk mengklasifikasikan orang yang mempunyai penyakit jantung atau tidak dengan beberapa tahapan pemeriksaan medis. Evaluasi terhadap data yang digunakan pada penelitian ini adalah menggunakan confusion matrix. Penelitian dilakukan dengan menggunakan variabel penelitian sebanyak 13 dan jumlah data sebanyak 1025. Data tersebut kemudian dibagi menjadi data *training* dan *testing* dengan perbandingan 80:20. Menggunakan metode *K-Nearest Neighbour* dapat diperoleh nilai K yang menunjukkan akurasi terbaik dalam pengklasifikasian. Nilai K yang menunjukkan evaluasi terbaik yang diperoleh dalam penelitian ini adalah nilai K = 2 dengan akurasi 89 %, sensitifitas (*recall*) 82 %, dan spesifitas 97 %.

Kata Kunci: Penyakit Jantung, Klasifikasi, Algoritma *K-Nearest Neighbour*.

Abstract

The heart is the most important organ owned by humans. If there is damage caused by narrowing of the blood vessels, it will cause heart disease. Heart disease ranks as the leading cause of death in Indonesia. This study aims to classify people who have heart disease or not with several stages of medical examination. Evaluation of the data used in this study is using confusion matrix. The research was conducted using 13 research variables and a total of 1025 data. The data is then divided into training and testing data with a ratio of 80:20. Using the *K-Nearest Neighbor* method, the K value can be obtained which shows the best accuracy in classification. The K value that shows the best evaluation obtained in this study is the value of K = 2 with an accuracy of 89%, sensitivity (*recall*) 82%, and specificity 97%.

Keywords: Heart Disease, Classification, *K-Nearest Neighbour* Algorithm

Pendahuluan

Jantung merupakan organ yang paling penting pada tubuh manusia, yang mempunyai fungsi untuk mengalirkan darah yang membawa oksigen keseluruh tubuh. Apabila jantung terjadi kerusakan maka akan menyebabkan penyakit jantung[1]. Penyakit jantung merupakan kejadian dimana suatu bagian jantung mengalami penyumbatan pembuluh darah. Penyakit jantung berkaitan erat dengan faktor resiko seperti hipertensi, kolesterol dalam darah, gula darah, asam urat, umur, *life style* dan lain-lain. Identifikasi kerentanan penyakit jantung dilaksanakan dengan berdasarkan data faktor-faktor resiko seseorang terhadap penyakit jantung[2].

Organisasi Kesehatan dunia WHO tahun 2019 memberitakan kematian yang disebabkan oleh penyakit jantung adalah sebanyak 31% dari kematian di seluruh dunia[3]. Dikutip oleh *Global Burden of Disease* dan IHME 2014-2019, penyakit jantung juga adalah penyebab kematian utama di Indonesia. Data ini dari hasil riset Kesehatan (Riskesdas) 2013 dan 2018. Terjadi peningkatan penyakit jantung, dari 0,5% pada tahun 2013 menjadi 1,5% pada tahun 2018[4].

Tingginya angka kematian akibat penyakit jantung yang menjadi alasan penelitian ini dilakukan. Seiringnya berjalan waktu teknologi juga semakin pesat. Memudahkan kita dalam kegiatan medis untuk mengetahui mempunyai penyakit jantung atau tidak. Salah satu cara dari kemajuan teknologi yang memudahkan dalam kegiatan medis adalah metode klasifikasi. Klasifikasi merupakan bentuk metode atau teknik yang digunakan untuk mengumpulkan beberapa objek menjadi beberapa kelompok. Salah satu metode dalam klasifikasi yaitu algoritma *K-Nearest Neighbour* adalah algoritma yang menggunakan konsep dasar yaitu jarak antar data dengan data lain agar berdekatan[5].

Dari penelitian sebelumnya yang menggunakan KNN juga menyelesaikan masalah penyakit jantung dengan klasifikasi KNN. Seperti penelitian *K-Nearest Neighbour* pada penyakit ginjal kronis [6] dengan tingkat akurasi klasifikasi yang cukup tinggi sebesar 93,44%. Terdapat pula penelitian sebelumnya tentang klasifikasi *K-Nearest Neighbour* Dan Gabor filter serta Wiener filter untuk mendiagnosis penyakit pneumonia [7] yang mempunyai nilai akurasi paling tinggi yaitu dengan nilai $K = 5$ sebesar 79,65% dan nilai akurasi terendah dengan nilai $K = 7$ sebesar 71,28%. Dari beberapa penelitian sebelumnya terbukti bahwa klasifikasi *K-Nearest Neighbour* yang mempunyai akurasi klasifikasi dengan nilai yang tinggi. Sehingga pada penelitian ini digunakan penerapan Algoritma *K-Nearest Neighbour* untuk mengetahui tingkat akurasi yang akan diperoleh dari data penyakit jantung.

Landasan Teori

Berikut ini adalah beberapa teori yang digunakan sebagai penunjang penelitian

1. Penyakit Jantung
Penyakit jantung mempunyai istilah yang umum untuk apa saja jenis kelainan yang menyerang pada jantung. Penyakit jantung mempunyai arti yang sama dengan penyakit kardiovaskular, namun bukan merupakan penyakit kardiovaskular. Penyakit kardiovaskular mengacu pada 26,6 juta orang yang didiagnosis (11,3% dari populasi orang dewasa) [8].
2. Data Mining
Data mining adalah suatu proses analitis untuk menemukan pengetahuan dalam kumpulan data yang kompleks dan besar. Data mining adalah teknik yang terletak di persimpangan antara statistik dan ilmu komputer. Lebih tepatnya, data mining merupakan hasil kombinasi statistik, ilmu komputer, kecerdasan buatan, dan pembelajaran mesin[9].
3. Normalisasi Data
Normalisasi data adalah rumus statistik yang memodifikasi kumpulan data, sehingga mempunyai variasinya antara 1 dan 0. Hal ini dapat berguna untuk menyamakan variansinya dari -1 sampai 1. Normalisasi data dapat juga membantu perbandingan antara nilai normalisasi yang sesuai dari dua atau beberapa banyak data yang berbeda dengan menghilangkan pengaruh variasi dalam penskalaan data. Metode normalisasi data yang paling sering dipakai adalah sistem Min-Max[10].

$$X_{new} = \frac{X_{old} - X_{min}}{X_{max} - X_{min}}$$

Dimana:

X_{new} = Hasil normalisasi nilai data tersebut

X_{old} = Nilai data yang lama
 X_{min} = Nilai data terkecil dalam satu data
 X_{max} = Nilai data terbesar dalam satu data

4. Klasifikasi

Klasifikasi merupakan suatu model yang digunakan untuk mengumpulkan data menjadi beberapa kelas dengan tujuan agar data data yang masih tidak diketahui kelasnya dapat diklasifikasikan sesuai kelasnya[11]. Klasifikasi adalah contoh tugas pembelajaran mesin yang diawasi (*mechine learning*) untuk membuat prediksi. Pada klasifikasi juga, kita dapat mengidentifikasi orang atau objek dalam kategori tertentu. Informasi mengenai objek atau data sebelumnya dapat digunakan oleh algoritma sebagai bahan untuk memperoleh aturan[12].

Klasifikasi data terbagi menjadi 2 yaitu data training dan data testing. Data training adalah data yang berisi nilai dari data training dan data testing yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor. Data testing adalah data baru yang akan diklasifikasikan sesuai dengan model yang telah dibuat dan memilih dengan akurasi klasifikasinya yang tinggi[13].

5. Algoritma K-Nearest Neighbour

Algoritma K-Nearest Neighbour atau yang biasa kita ketahui dengan KNN adalah suatu algoritma non-parametric. Algoritma K-Nearest Neighbour juga merupakan salah satu algoritma yang paling banyak dan sering digunakan untuk memprediksi dan klasifikasi[14]. K-Nearest Neighbor (KNN) termasuk dalam kelompok pembelajaran berbasis individu. Algoritma ini juga merupakan teknik belajar malas atau yang disebut juga sebagai Teknik *lazy learning*. KNN dilakukan dengan mencari nilai k kelompok objek pada data latih yang terdekat (serupa) dengan objek pada data baru atau data uji[15].

Dalam menentukan k tetangga terdekat pada suatu titik data, perlu dilakukan pengukuran perbedaan antar titik data. Titik – titik data tersebut dapat diukur dengan beberapa Teknik, misalnya yaitu jarak Euclidian, jarak Minkowski, jarak Hamming, koefisien korelasi Pearson dan persamaan cosine. Dalam penelitian kali ini untuk menentukan k digunakan rumus Euclidian distance seperti dibawah ini:

$$d_i = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

Dimana:

p : Dimensi data
 i : Atribut data
 x : Data sampel / data latih
 y : Data uji / data testing
 d : Jarak

Tahapan dalam K-Nearest Neighbor[16]:

- 1) Menentukan nilai K (jumlah tetangga terdekat). nilai K pada saat pengujian ditentukan berdasarkan nilai K optimal pada saat pelatihan.
- 2) Hitung jarak Euclidean setiap fitur pada sampel data.
- 3) Mengurutkan benda-benda tersebut ke dalam kelompok-kelompok dengan jarak Euclidean terkecil.
- 4) Mengumpulkan data testing (klasifikasi tetangga terdekat)
- 5) Dengan memasukkan data testing maka dapat diperoleh hasil klasifikasi.

6. Confusion Matrix

Confusion matrix adalah metode yang umum digunakan untuk menghitung akurasi algoritma K-Nearest Neighbour. *Confusion matrix* merupakan cara untuk menganalisis seberapa baik

pengklasifikasi dapat mengenali tupel dari kelas yang berbeda[17]. Dari *Confusion matrix* dapat dihitung jumlah sensitifitas (*recall*), spesifitas, akurasi, precision.

Tabel 1. *Confusion Matrix*

<i>Correct Classification</i>	<i>Classification</i>	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Untuk menghitung nilai sensitifitas (*recall*), spesifitas, akurasi dan precision menggunakan persamaan dibawah ini:

a) Sensitifitas

Untuk mengetahui nilai sensitifitas yaitu membandingkan jumlah TP dengan jumlah record positif. Rumusnya sebagai berikut:

$$\text{Sensitifitas} = \frac{TP}{P}$$

b) Spesifitas

Untuk mengetahui nilai spesifitas yaitu membandingkan jumlah TN dengan jumlah record negatif. Rumusnya sebagai berikut:

$$\text{Spesifitas} = \frac{TN}{N}$$

c) Akurasi

Untuk mengetahui nilai akurasi yaitu dengan cara membandingkan kasus yang benar dengan jumlah seluruh kasus. Rumusnya sebagai berikut:

$$\text{Akurasi} = \frac{TN + TP}{TN + TP + FP + FN}$$

d) Presisi

Presisi merupakan nilai untuk menunjukkan tingkat keberhasilan sebuah model dalam mengenali suatu objek. Rumusnya sebagai berikut:

$$\text{Presisi} = \frac{TP}{TP + FP}$$

Keterangan[18]:

- TP* = Jumlah benar positif
- TN* = Jumlah benar negatif
- FP* = Jumlah salah positif
- FN* = Jumlah salah negative
- P* = Jumlah record positif
- N* = Jumlah tupel negative

Bahan dan Metode

Pada penelitian ini data yang digunakan adalah data penyakit jantung yang diperoleh dari situs Kaggle. *Data mining* yang berisi beberapa tahapan pemeriksaan seseorang tersebut mempunyai penyakit jantung atau tidak dengan ouput 1 adalah mempunyai penyakit jantung dan jika output 0 adalah tidak mempunyai

penyakit jantung. Sementara variabel yang digunakan dalam penelitian ini terdapat 14 variabel dan 1025 data. Terdapat 13 variabel yang menjadi variabel independent (X) yaitu terdiri dari beberapa pemeriksaan yang dilakukan dan satu variabel dependen (Y) yaitu nilai target dengan angka 1 dan 0. Berikut variabel independent dan variabel dependen yang digunakan dalam klasifikasi algoritma *K-Nearest Neighbour*:

Tabel 2. Sampel Data Penelitian

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	Y
52	1	0	125	212	0	1	168	0	1	2	2	3	0
53	1	0	140	203	1	0	155	1	3,1	0	0	3	0
70	1	0	145	174	0	1	125	1	2,6	0	0	3	0
61	1	0	148	203	0	1	161	0	0	2	1	3	0
62	0	0	138	294	1	1	106	0	1,9	1	3	2	0
...
47	1	0	110	275	0	0	118	1	1	1	1	2	0
50	0	0	110	254	0	0	159	0	0	2	0	2	1
54	1	0	120	188	0	1	113	0	1,4	1	1	3	0

Keterangan:

Y : Target

X1 : Usia

X2 : Seks

X3 : Jenis nyeri dada (4 nilai)

X4 : Tekanan darah istirahat

X5 : Kolesterol dalam mg/dl

X6 : Gula darah >120 mg/dl

X7 : Hasil elektrokardiografi (nilai 0,1,2)

X8 : Detak jantung maksimum tercapai

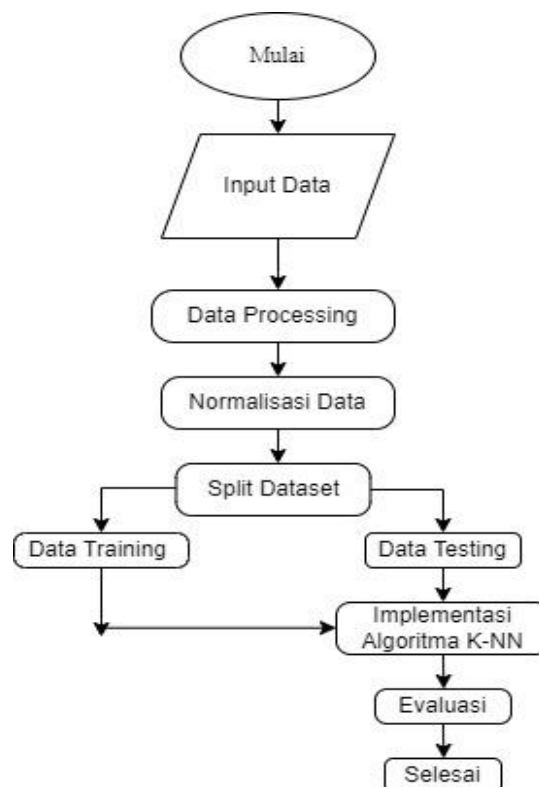
X9 : Angina akibat olahraga

X10 : Oldpeak

X11 : Kemiringan puncak latihan segmen St

X12 : Jumlah pembuluh besar (0-3)

X13 : thal,



Gambar 1. Flowchart

Hasil dan Pembahasan

Pada penelitian ini yang terdiri dari 13 variabel independent (X) dan varibel dependen (Y) dengan jumlah data 1025 dilakukan pengolahan data terlebih dahulu. Pada data penyakit jantung terdapat variabel X10 yaitu oldpeak yang bukan integer sehingga dilakukannya perubahan agar variabel X10 menjadi integer dan hasilnya terdapat pada tabel dibawah ini.

Tabel 3. Hasil Data yang telah diubah

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	Y
52	1	0	125	212	0	1	168	0	1	2	2	3	0
53	1	0	140	203	1	0	155	1	3	0	0	3	0
70	1	0	145	174	0	1	125	1	2	0	0	3	0
61	1	0	148	203	0	1	161	0	0	2	1	3	0
62	0	0	138	294	1	1	106	0	1	1	3	2	0
...
47	1	0	110	275	0	0	118	1	1	1	1	2	0
50	0	0	110	254	0	0	159	0	0	2	0	2	1
54	1	0	120	188	0	1	113	0	1	1	1	3	0

Kemudian setelah mengolah data dilakukannya normalisasi data dengan menggunakan Min – Max dengan rentang -1 sampai 1. Dilakukannya normalisasi data agar setiap data tidak ada jarak atau skala yang terlalu besar pada data tersebut.

Tabel 4. Hasil Data Normalisasi

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
0.48	1.0	0.00	0.29	0.20	0.0	0.5	0.74	0.0	0.17	1.0	0.50	1.00
0.50	1.0	0.00	0.43	0.17	1.0	0.0	0.64	1.0	0.50	0.0	0.00	1.00
0.85	1.0	0.00	0.48	0.11	0.0	0.5	0.41	1.0	0.33	0.0	0.00	1.00
0.67	1.0	0.00	0.51	0.17	0.0	0.5	0.69	0.0	0.00	1.0	0.25	1.00
0.69	0.0	0.00	0.41	0.38	1.0	0.5	0.27	0.0	0.17	0.5	0.75	0.67
...
0.37	1.0	0.00	0.15	0.34	0.0	0.0	0.36	1.0	0.17	0.5	0.25	0.67
0.44	0.0	0.00	0.15	0.29	0.0	0.0	0.67	0.0	0.00	1.0	0.00	0.67
0.52	1.0	0.00	0.24	0.14	0.0	0.5	0.32	0.0	0.17	0.5	0.25	1.00

Setelah tahap normalisasi selesai kemudian bagi data menjadi 2 bagian yaitu 80% data training dan 20% data testing. Setelah terbaginya data, untuk mengetahui orang tersebut mempunyai penyakit jantung atau tidak adalah melakukan pengujian dengan menggunakan data testing.

Tabel 5. Data Testing

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	Y
52	1	0	125	212	0	1	168	0	1	2	2	3	0
53	1	0	140	203	1	0	155	1	3	0	0	3	0
70	1	0	145	174	0	1	125	1	2	0	0	3	0
61	1	0	148	203	0	1	161	0	0	2	1	3	0
62	0	0	138	294	1	1	106	0	1	1	3	2	0
...
70	1	1	156	245	0	0	143	0	0	2	0	2	1
45	1	0	115	260	0	0	185	0	0	2	0	2	1
42	1	3	148	244	0	0	178	0	0	2	2	2	1

Tabel 6. Data Training

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	Y
58	0	0	170	225	1	0	146	1	2	1	2	1	0
61	1	0	140	207	0	0	138	1	1	2	1	3	0
62	0	0	140	268	0	0	160	0	3	0	2	2	0
60	1	0	130	253	0	1	144	1	1	2	1	3	0
54	1	0	140	239	0	1	160	0	1	2	0	2	1
...
47	1	0	110	275	0	0	118	1	1	1	1	2	0
50	0	0	110	254	0	0	159	0	0	2	0	2	1
54	1	0	120	188	0	1	113	0	1	1	1	3	0

Penerapan algoritma K- *Nearest Neighbour* yang dilakukan setelah membagi data menjadi data 80 persen data *training* dan 20 data *testing*. Diperhatikan bahwa nilai K yang terbaik ada diantara K = 2, K = 4 dan K = 6, sebagaimana tabel berikut

Tabel 7. Hasil Akurasi Nilai K

	K= 2	K=4	K=6
Akurasi	89%	77%	77%

Terlihat pada plot di atas dari nilai K 1 – 10 yang mempunyai nilai akurasi terbaik adalah 1 dan 2. Pada penelitian ini menggunakan nilai K yaitu 2 dengan akurasi 89 %. Karena ada dari penelitian terdahulu yang mengatakan bahwa nilai K itu tidak boleh 1. Setelah mendapatkan nilai K, kemudian mencari nilai Y prediksi dari data testing yang digunakan untuk mengetahui orang tersebut mempunyai penyakit jantung atau tidak.

Tabel 8. Data Y Prediksi

Y prediksi
1
1
0
0
0
...
...
0
0
0
1

Akurasi dari nilai K yang diatas diperoleh dari *confusion matrix*. *Confusion matrix* ini bertujuan untuk mengetahui tingkat akurasi dari pengklasifikasian pada penelitian ini.

Tabel 9. Hasil *Confusion Matrix*

	Mempunyai Penyakit Jantung	Tidak Mempunyai Penyakit Jantung
Mempunyai Penyakit Jantung	101	3
Tidak Mempunyai Penyakit Jantung	18	83

Berdasarkan *confusion matrix* di atas dapat dilakukan evaluasi terhadap pengklasifikasian pada penelitian ini. Evaluasi yang digunakan adalah akurasi, sensitifitas (*recall*), dan spesitifitas.

Tabel 10. Hasil Evaluasi

Akurasi	Presisi	Recall	Spesitifitas
89%	96%	82%	97%

Berdasarkan tabel di atas didapatkan akurasi sebesar 89 %, sensitifitas (*recall*) sebesar 82 % dan spesifitas sebesar 97 %.

Kesimpulan

Berdasarkan dari beberapa tahapan penelitian yang telah digunakan pada implementasi Algoritma K-Nearest Neighbour untuk mengetahui orang tersebut mempunyai penyakit jantung atau tidak. Kesimpulan yang didapatkan yaitu implementasi yang dihitung menggunakan *google colab*, kemudian menormalisasi data agar setiap data tidak mempunyai jarak yang terlalu besar. Kemudian dibagi menjadi 80:20 dengan 80 persen data *training* dan 20 persen data *testing*. Penentuan K dengan membandingkan nilai K=2, K= 4 dan K=6, dan kemudian diperoleh nilai K=2 yang merupakan nilai K dengan akurasi tertinggi yaitu sebesar 89%. Dengan demikian evaluasi terhadap pengklasifikasian juga tinggi yaitu dengan akurasi 89%, sensitifitas (*recall*) sebesar 82% dan spesifitas sebesar 97%.

Referensi

- [1] L. A. Dewi, "Klasifikasi Machine Learning Untuk Mendeteksi Penyakit Jantung Dengan Algoritma K-NN, Decision Tree dan Random Forest," *Repository.Uinjkt.Ac.Id*, 2023, [Online]. Available: [https://repository.uinjkt.ac.id/dspace/handle/123456789/71124%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/71124/1/LIZKY ASKA DEWI-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/71124%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/71124/1/LIZKY%20ASKA%20DEWI-FST.pdf)
- [2] W. Lestari and S. Sumarlinda, "Studi Komparatif Model Klasifikasi Kerentanan Penyakit Jantung Menggunakan Algoritma Machine Learning," *SATIN - Sains dan Teknol. Inf.*, vol. 9, no. 1, pp. 107–115, 2023, doi: 10.33372/stn.v9i1.918.
- [3] WHO, "WHO Health Statistics Overview 2019," 2021.
- [4] Kementerian Kesehatan Republik Indonesia, "Penyakit Jantung Penyebab Utama Kematian, Kemenkes Perkuat Layanan Primer," *Public*, no. September 2022. Jakarta, 29 September 2022, pp. 1–1, 2022. [Online]. Available: <https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20220929/0541166/penyakit-jantung-penyebab-utama-kematian-kemenkes-perkuat-layanan-primer/>
- [5] M. F. Akbarollah, W. Wiyanto, D. Ardiatma, and A. T. Zy, "Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Jantung," *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 850–860, 2023, doi: 10.47065/josyc.v4i4.4071.
- [6] A. Ariani and Samsuryadi, "Klasifikasi Penyakit Ginjal Kronis menggunakan K-Nearest Neighbor," *Pros. Annu. Res. Semin. 2019*, vol. 5, no. 1, pp. 148–151, 2019.
- [7] F. Antony, H. Irsyad, and M. E. Al Rivan, "KNN Dan Gabor Filter Serta Wiener Filter Untuk Mendiagnosis Penyakit Pneumonia Citra X-RAY Pada Paru-Paru," *J. Algoritm.*, vol. 1, no. 2, pp. 147–155, 2021, doi: 10.35957/algoritm.v1i2.893.
- [8] A. Riski, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *J. Tek. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: <https://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141/156>
- [9] C. Armayani, A. Fauzi, and H. Sembiring, "Implementasi Data Mining Pengelompokan Jumlah Data Produktivitas Ubinan Tanaman Pangan Berdasarkan Jenis Ubinan Dengan Metode Clustering Dikab Langkat (Studi Kasus : Badan Pusat Statistik Langkat)," *J. Inform. Kaputama*, vol. 5, no. 1, pp. 185–196, 2021, doi: 10.59697/jik.v5i1.318.
- [10] D. A. Ryfai, N. Hidayat, and E. Santoso, "Klasifikasi Tingkat Resiko Serangan Penyakit Jantung menggunakan Metode K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 10, pp. 4701–4707, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [11] M. Kamber and J. Han, *Data Mining: Concepts and Techniques : Concepts and Techniques*. 2018.
- [12] I. R. Amilia, H. Oktavianto, and G. Abdurrahman, "Penerapan Backward Elimination Untuk Seleksi Fitur Pada Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Gagal Jantung," *J. Smart Teknol.*, vol. 1, no. 1, pp. 2774–1702, 2021, [Online].

Available: <http://jurnal.unmuhjember.ac.id/index.php/JST>

- [13] A. J. Nathan and A. Scobell, "Model Algoritma K-nearest Neighbor untuk memprediksi kelulusan mahasiswa," *Foreign Aff.*, vol. 91, no. 5, pp. 1–9, 2012.
- [14] T. A. Munandar and A. Q. Munir, "Implementasi K-Nearest Neighbor Untuk Prototype Sistem Pakar Identifikasi Dini Penyakit Jantung," *Respati*, vol. 17, no. 2, p. 44, 2022, doi: 10.35842/jtir.v17i2.457.
- [15] S. Sahar, "Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Classifier Pada Dataset Penyakit Jantung," *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 79–86, 2020, doi: 10.33096/ijodas.v1i3.20.
- [16] A. Yandi Saputra and Y. Primadasa, "Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbour Implementation of Classification Method to Predict Student Graduation Using K-Nearest Neighbor Algorithm," *Techno.Com*, vol. 17, no. 4, p. 9, 2018.
- [17] A. Sumiah and N. Mirantika, "Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan," *Buffer Inform.*, vol. 6, no. 1, pp. 1–10, 2020.
- [18] H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *J. Penelit. Ilmu Komputer, Syst. Embed. Log.*, vol. 1, no. 1, pp. 65–76, 2013.