

Prediksi Resiko Penyakit Jantung dan Pembuluh Darah Menggunakan Algoritma K-Nearest Neighbor (KNN)

Heri Susanto✉, Dwi Yanto, Ninanesia Rusdiana, Deasy Tri Rahayu

Prodi Sistem Informasi, Akademi Manajemen Informatika dan Komputer Taruna, Probolinggo, Indonesia

Info Articles

Keywords:
Cardiovascular disease, Data Mining, K-Nearest Neighbors (KNN), Risk Assessment

Abstrak

Penyakit jantung adalah salah satu penyakit paling umum yang menjadi penyebab kematian seseorang di seluruh dunia. Di Indonesia, pada tahun 2014, survei Sample Registration System (SRS) menjelaskan bahwa 12,9% pelaku utama terjadinya kematian pada semua umur adalah penyakit jantung jenis koroner. Pendeteksian dini kemungkinan terserang penyakit jantung dan pembuluh darah sangat diperlukan guna mencegah kemungkinan terburuk yang bisa terjadi pada setiap orang. Penelitian ini bertujuan untuk memberikan gambaran tingkat resiko penyakit jantung dan pembuluh darah pada UPT Puskesmas Sumberasih. Suatu metode klasifikasi yang dapat diimplementasikan ke sebuah perangkat lunak untuk mendeteksi tingkat resiko penyakit jantung dan pembuluh darah adalah K-Nearest Neighbors (KNN), yakni suatu metode yang melakukan klasifikasi berdasarkan data latih dengan cara menghitung jarak yang paling dekat dengan rumus *Euclidian Distance* terhadap obyek berkelas sebanyak K. Pada penelitian ini, menggunakan dataset sebanyak 200, pengujian menggunakan metode *split validation* dimana jumlah data latih sebanyak 180, dan data uji sebanyak 20, dengan parameter nilai K=3 didapatkan akurasi sebesar 95%.

Abstract

Cardiovascular disease remains one of the leading causes of mortality worldwide. In Indonesia, a 2014 Sample Registration System (SRS) survey reported that coronary heart disease accounted for 12.9% of all deaths across all age groups. Early detection of the risk of cardiovascular disease is crucial in preventing potentially adverse outcomes. This study aims to assess the risk of cardiovascular disease and vascular conditions among patients at the Sumberasih Community Health Center (UPT Puskesmas). The study employs the K-Nearest Neighbors (KNN) classification method, which can be implemented in software to predict the risk of cardiovascular disease and vascular conditions. KNN classifies data based on the proximity of the data points, utilizing the Euclidean Distance formula to determine the K-nearest neighbors for a given data point. Through analysis and testing, this research indicates that the accuracy of risk prediction is influenced by the quantity of training data. The highest accuracy, 95%, was achieved

when utilizing 200 training data samples, setting $K=3$, and testing with 20 data samples.

This research provides valuable insights into the potential for using KNN classification as a tool for the early detection of cardiovascular disease and vascular conditions, which can significantly contribute to better healthcare and disease prevention strategies.

PENDAHULUAN

Penyakit jantung adalah suatu peristiwa dimana organ jantung mengalami gangguan dalam melakukan fungsinya dalam melakukan siklus aliran darah di tubuh. Ada beberapa jenis gangguan, seperti gangguan pada pembuluh darah jantung, penyempitan atau penyumbatan pembuluh darah, katup jantung, irama jantung, dan sebagainya. Jantung dikatakan berjalan dengan semestinya apabila darah dapat mengalir keluar dan masuk di jantung dengan lancar (Ryfai et al., 2022).

Salah satu masalah utama kesehatan di Indonesia adalah Penyakit Jantung dan pembuluh darah. Riset kesehatan dasar badan penelitian dan pengembangan kesehatan kementerian kesehatan tahun 2013 menunjukkan prevalensi penyakit jantung di Indonesia 7,2 % atau 7 dari 100 orang menderita penyakit jantung, sementara prevalensi Stroke sebesar 12,1% atau 12 dari 1000 orang menderita penyakit stroke (Hasran, 2020).

Survey sample Registration system (SRS) Indonesia tahun 2014, Stroke menjadi penyebab kematian tertinggi pada semua umur sebesar 21,1%, penyakit jantung coroner sebanyak 12,9% (Indriasih et al., 2020; Sinambela L.P., 2014). Dibanding dengan penyakit katastropik lainnya, penyakit jantung menduduki peringkat teratas untuk biaya rawat inap. Data Badan penyelenggara jaminan social bidang Kesehatan (BPJS) tahun 2015 menunjukkan biaya rawat inap total 6,9 Triliun rupiah. Penyakit jantung juga dapat ditemukan pada penduduk usia muda, penyakit jantung dan pembuluh darah sebenarnya dapat dicegah dengan mengurangi faktor resiko merokok, diet tidak sehat, kurang aktifitas fisik dan konsumsi alkohol .

WHO bekerja sama dengan berhasil menyusun Carta prediksi resiko. Salah satu instrument untuk menilai resiko penyakit jantung dan pembuluh darah dalam 10 tahun kedepan. Carta prediksi resiko bekerja dengan memprediksi resiko mengalami kejadian jantung dan pembuluh darah berdasarkan ada tidaknya diabetes mellitus, jenis kelamin, status merokok, usia > 40 tahun, tekanan darah sistolik, kadar kolesterol(Santoso et al., 2023).

Carta prediksi resiko dapat digunakan untuk memotivasi penderita terutama untuk merubah gaya hidup, mengkonsumsi obat anti hipertensi, obat penurun lemak darah serta asam asetyl calisilat.

Data mining merupakan proses menemukan pola dan pengetahuan dari data yang berjumlah banyak. Sumber data adalah hal mendasar yang harus tersedia untuk dilakukan proses data mining(Yanto et al., 2023).

Dengan semakin berkembang dan meningkatnya kemampuan perangkat komputer, pemrosesan data mining semakin bisa ditingkatkan (Susanto & Jamal, 2020).

Hasran (2020), melakukan penelitian dalam mengimplementasikan metode k-NN untuk klasifikasi penyakit jantung. Pada penelitian ini mencakup pengukuran performa (akurasi, presisi, recall dan f-measure) metode KNN dengan nilai K=3 hingga K=9 pada objek 1000 data pasien penyakit jantung yang diperoleh dari pusat dataset UCI machine Learning Repository. Hasil dari pengukuran performa diperoleh nilai K terbaik adalah 6 dimana nilai akurasi 85%, presisi 78%, recall 93% dan f-measure sebesar 85%.

Ahmad Denis Ryfai, Nurul Hidayat, Edy Santoso (2022), melakukan penelitian dalam mengimplementasikan metode k-NN untuk Klasifikasi tingkat resiko serangan jantung. Pengujian yang dilakukan terdapat 2 jenis, yaitu pengujian pertama untuk meneliti akurasi dari pengaruh jumlah data latih dan pengujian kedua untuk meneliti akurasi dari pengaruh nilai K. Data yang digunakan dalam pengujian berupa data uji sebanyak 60, data latih, dan nilai K. Pada pengujian pertama dalam mengetahui pengaruh jumlah data latih, didapatkan akurasi tertinggi sebesar 88,333% yang dihasilkan ketika data latih sebanyak 126 dengan K bernilai 5. Pada pengujian kedua dalam mengetahui pengaruh nilai K, didapatkan akurasi tertinggi sebesar 96,667% (Ryfai et al., 2022).

Happy Andrian Dwi Fasnuari, Haris Yuana, M. Taofik Chulkamdi (2022), menerapkan Algoritma K-Nearest Neighbor (KNN) dalam menentukan klasifikasi penyakit diabetes melitus. Penelitian ini menggunakan 81 data training dan 54 data testing yang sudah mengalami proses pembersihan data kemudian dilakukan proses normalisasi. Pengujian Algoritma K-Nearest Neighbor (KNN) dalam menentukan klasifikasi penyakit diabetes melitus menggunakan confusion matrix mendapatkan hasil akurasi yang baik sebesar 93%, presisi sebesar 100%, Recall sebesar 60% dan F1-Score sebesar 75% (Dwi Fasnuari et al., 2022).

Dari sumber data yang didapatkan dari UPT Puskesmas Sumberasih dalam 3 tahun terakhir pasien yang dirujuk karna penyakit jantung berjumlah 304 pasien, diantaranya laki-laki berjumlah 151 orang dan perempuan berjumlah 153 orang (Sumberasih, 2022).

Dari data tersebut, selanjutnya diolah dan dilaporkan kepada pihak eksekutif untuk mengetahui perkembangan puskesmas pada waktu tertentu.

Dari permasalahan di atas untuk mengatasi banyaknya penyakit jantung yang ditimbulkan, maka dilakukannya analisis dan prediksi menggunakan Algoritma K-Nearest Neighbors. Yang mana nantinya akan menentukan jenis penyakit jantung yang sering terjadi atau sering di alami, hasil dari prediksi tersebut akan dibandingkan jenis penyakit jantung yang paling banyak dialami, hasil tersebut nantinya bisa dilihat dari tabel atau grafik prediksi yang sudah di analisis.

METODE

1. Pengumpulan Data

Dari hasil observasi di UPT Puskesmas Sumberasih Kabupaten Probolinggo, didapatkan data rekam medis pasien dengan jangka waktu selama tahun 2022. Jumlah data tiap bulannya bervariasi, tergantung dari jumlah kedatangan pasien tiap pekan. Sumber data dari register harian jumlah kunjungan pasien selama tahun 2022 di UPT Puskesmas Sumberasih sebanyak 21.865 pasien. Adapun data penelitian ini, digunakan sebanyak 200 data pasien dengan pemeriksaan penunjang laboratorium yaitu pemeriksaan kolesterol yang diambil dari bulan Januari 2022 – Desember 2022 (Sumberasih, 2022).

2. Pengolahan Data Awal

Pada tahap pendeteksian tingkat resiko penyakit jantung dengan kecerdasan buatan, diperlukan kumpulan data pasien beserta dengan fitur yang berhubungan dengan penyakit tersebut, seperti faktor fisik tubuh, gejala penyakit, atau hasil kadar kolesterol. Adapun atribut yang digunakan adalah hasil pemeriksaan pasien serta hasil pemeriksaan laboratorium (kolesterol). Selanjutnya, dari atribut yang telah ditetapkan akan menghasilkan label / target berupa tingkat resiko pada penyakit jantung dan pembuluh darah.

Data yang diperoleh selanjutnya diolah dan disesuaikan dengan kebutuhan penelitian. Data tersebut dikategorikan untuk memudahkan proses pengujian metode. Data pasien dikategorikan berdasarkan parameter/atribut yang telah ditentukan yaitu jenis kelamin, umur, riwayat penyakit Diabetes Melitus, merokok, tekanan darah systole, dan hasil pemeriksaan kolesterol. Parameter/atribut tersebut menjadi variable input yang diuraikan / dijadikan numerik sebagai berikut :

- x1 = jenis kelamin (nilai diubah 1 untuk “laki-laki” dan 0 untuk “perempuan”)
- x2 = umur pasien
- x3 = riwayat penyakit Diabetes Melitus (nilai diubah 1 untuk “ya” dan 0 untuk “tidak”)
- x4 = merokok (nilai diubah 1 untuk “ya” dan 0 untuk “tidak”)
- x5 = hasil tekanan darah sistole
- x6 = hasil pemeriksaan kolesterol

Selanjutnya parameter/atribut tersebut menghasilkan variabel output atau target atau label yaitu : y1 = Rendah, Sedang dan Tinggi

Semua parameter tersebut akan digunakan sebagai input dari proses training dan testing.

Tabel 1. Hasil Pengkategorian Data Pasien

No	Jenis Kelamin	Umur	Riwayat DM	Merokok	Sistolik	Kolesterol	Tingkat Resiko
1	Laki-laki	55	Tidak	Ya	160	274	sedang
2	Laki-laki	45	Tidak	Ya	120	283	rendah
3	Laki-laki	52	Tidak	Ya	130	294	rendah
4	Laki-laki	54	Tidak	Ya	110	182	rendah
5	Laki-laki	59	Tidak	Ya	160	306	sedang
6	Laki-laki	54	Tidak	Ya	110	168	rendah
7	Laki-laki	72	Tidak	Ya	140	280	tinggi
8	Laki-laki	47	Tidak	Ya	160	294	sedang
9	Perempuan	53	Tidak	Tidak	130	209	rendah
...
200	Perempuan	43	Tidak	Tidak	130	219	rendah

Tabel 2. Konversi ke data numerik

No	Jenis Kelamin	Umur	Riwayat DM	Merokok	Sistolik	Kolesterol	Tingkat Resiko
1	1	55	0	1	160	274	sedang
2	1	45	0	1	120	283	rendah
3	1	52	0	1	130	294	rendah
4	1	54	0	1	110	182	rendah
5	1	59	0	1	160	306	sedang
6	1	54	0	1	110	168	rendah
7	1	72	0	1	140	280	tinggi
8	1	47	0	1	160	294	sedang
9	0	53	0	0	130	209	rendah
...
200	0	44	0	0	90	263	rendah

3. Implementasi menggunakan Algoritma KNN

Pengujian menggunakan Algoritma KNN dilakukan dengan menggunakan tools rapidminer. Dataset yang digunakan sebanyak 200, terdiri dari 180 data training dan 20 data testing yang dipilih secara acak.

4. Pengukuran Kinerja Klasifikator

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua set data dengan benar. Tetapi kinerja suatu sistem tidak bisa bekerja 100% benar. Oleh karena itu sistem klasifikasi harus diukur kinerjanya, cara mengukur kinerja klasifikasi yang digunakan adalah confusion matrik (Prasetyo, 2014). Untuk menghitung akurasi digunakan formula:

$$Akurasi = \frac{Jumlah\ Data\ yang\ diprediksi\ benar}{Jumlah\ prediksi\ yang\ dilakukan}$$

Di mana:

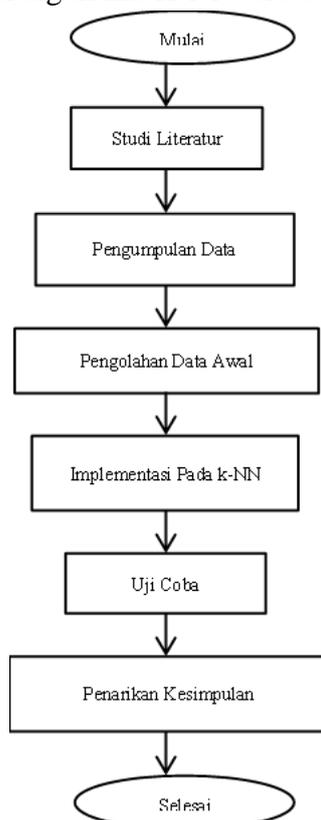
"Jumlah Data yang diprediksi benar" adalah jumlah data pengujian yang diklasifikasikan dengan benar oleh model KNN.

"Jumlah prediksi yang dilakukan" adalah total jumlah data yang diprediksi yang digunakan dalam pengujian model.

Semakin tinggi nilai akurasi, semakin baik model KNN Anda dalam memprediksi kelas data pengujian.

Pada ujicoba ini menggunakan parameter nilai $K=3$, $K=5$, dan $K=7$, selanjutnya hasil masing-masing dibandingkan untuk mendapatkan nilai akurasi tertinggi.

Langkah-langkah klasifikasi algoritma K-NN tersebut sebagai berikut:



Gambar 1. Langkah-langkah Klasifikasi dengan KNN

HASIL DAN PEMBAHASAN

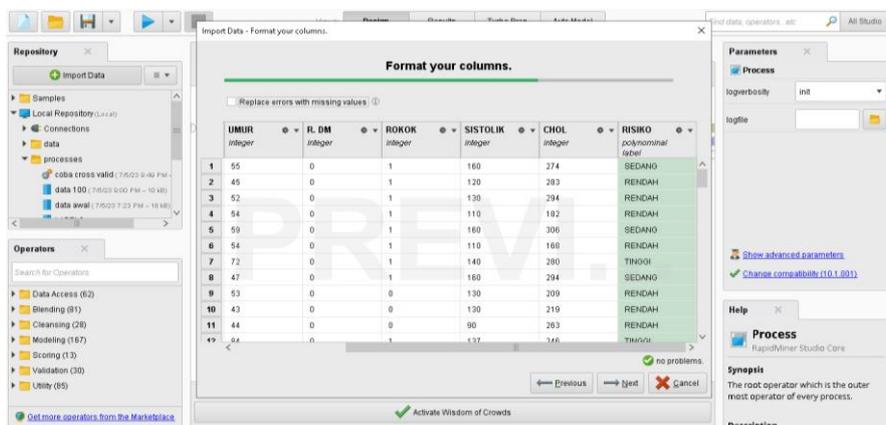
Pada penelitian ini, menerapkan data mining untuk melakukan prediksi penyakit jantung dan pembuluh darah menggunakan algoritma KNN berdasarkan pada 6 atribut, yakni : jenis kelamin, usia, riwayat DM, status merokok, tekanan darah sistolik, dan kolesterol. Dalam pemrosesan awal dilakukan data cleansing, data duplikat dan data yang tidak konsisten dihilangkan sehingga didapatkan dataset dengan total 200 record, dengan rincian 180 data training dan 20 data testing (Hermawati, 2013).

Tabel 3. Data Testing

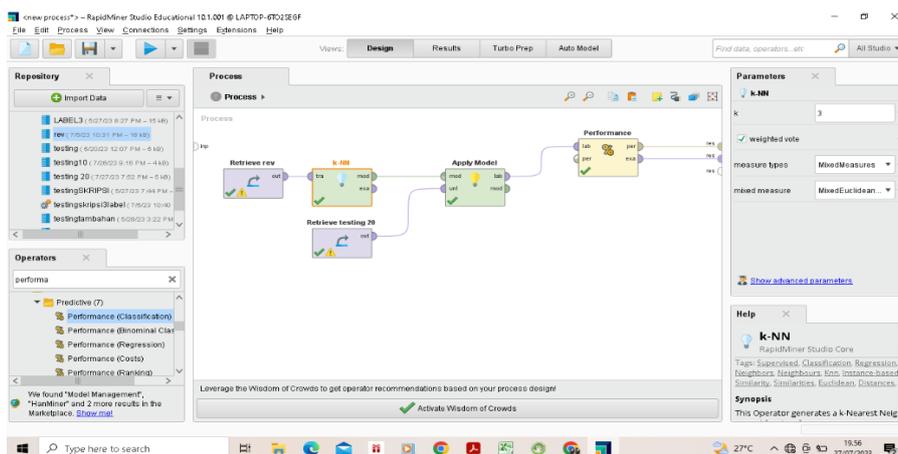
No	Jenis Kelamin	Umur	Riwayat DM	Merokok	Sistolik	Kolesterol	Tingkat Resiko
1	1	61	1	1	188	247	?

2	0	49	0	0	130	306	?
3	1	59	1	1	160	138	?
4	0	44	0	0	100	294	?
5	1	52	0	1	130	174	?
6	1	51	0	1	130	306	?
7	0	57	1	0	120	158	?
8	0	51	0	0	130	175	?
9	0	59	0	0	170	195	?
10	0	45	0	0	110	219	?
11	0	49	1	0	156	196	?
12	1	54	0	1	130	243	?
13	0	68	0	0	130	314	?
14	1	69	0	1	130	286	?
15	1	56	0	1	130	294	?
16	1	57	0	1	110	241	?
17	0	61	0	0	150	301	?
18	1	52	0	1	130	250	?
19	0	57	0	0	130	282	?
20	1	41	0	1	120	189	?

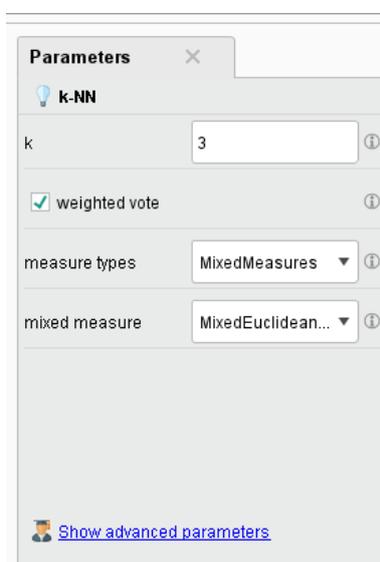
Pada penggunaan tools rapidminer, dataset yang sudah menjadi bentuk excel sheet di import ke dalam tools dengan memisahkan file excel antara data training dan data testing. Atur tabel dari excel tersebut menjadi komposisi atribut dan label sesuai dengan ketentuan yang telah dibuat sebelumnya



Gambar 2. Data Tarining pada Tools Rapid Miner



Gambar 3. Pemodelan Algoritma KNN pada Tools Rapid Miner



Gambar 4. Uji Coba parameter nilai K pada Tools Rapid Miner

Berikut ini adalah hasil melakukan uji coba terhadap nilai K:

1. Pengujian menggunakan parameter nilai K=3, menghasilkan akurasi sebesar 95%.

Tabel 4. Tabel Confusion Matrix parameter K=3

	True Tinggi	True Sedang	True Rendah	Class Precision
Prediksi Tinggi	1	0	0	100%
Prediksi Sedang	0	15	0	100%
Prediksi Rendah	0	1	3	75%

Sumber: Hasil proses menggunakan tools Rapid Miner.

2. Pengujian menggunakan parameter nilai K=5, menghasilkan akurasi sebesar 85%.

Tabel 5. Tabel Confusion Matrix parameter K=5

	True Tinggi	True Sedang	True Rendah	Class Precision
Prediksi Tinggi	0	0	0	0%

Prediksi Sedang	0	15	1	93,75%
Prediksi Rendah	1	1	2	50%

Sumber: Hasil proses menggunakan tools Rapid Miner.

3. Pengujian menggunakan parameter nilai $K=7$, menghasilkan akurasi sebesar 85%.

Tabel 6. Tabel Confusion Matrix parameter $K=7$

	True Tinggi	True Sedang	True Rendah	Class Precision
Prediksi Tinggi	0	0	0	0%
Prediksi Sedang	0	15	1	93,75%
Prediksi Rendah	1	1	2	50%

Sumber: Hasil proses menggunakan tools Rapid Miner.

Deteksi dinyatakan berhasil jika hasil dari penghitungan jarak menggunakan rumus euclidean distance dengan nilai k yang ditentukan dengan menggunakan cara manual dapat mendeteksi / adanya kecocokan diagnosa data testing dengan beberapa nilai k sesuai dengan diagnosa pada data testing awal.

Percobaan yang dilakukan dengan tiga nilai k berbeda menghasilkan prediksi dan nilai akurasi yang berbeda pula. Semakin besar nilai k tidak memberikan pengaruh terhadap semakin besarnya pula nilai akurasi yang didapat. Namun dari beberapa nilai k yang diuji coba, memiliki hasil yang berbeda dalam keberhasilan kasus prediksi/identifikasi benar terhadap resiko sedang dengan jumlah terbanyak dan nilai akurasi yang terbesar saat nilai $k = 3$.

SIMPULAN

Metode K-Nearest Neighbors (KNN) dapat digunakan untuk melakukan prediksi resiko mengalami kejadian penyakit jantung dan pembuluh darah, dimana diperlukan data latih, dan data uji, selanjutnya sistem akan melakukan perhitungan menggunakan metode KNN yang memiliki beberapa tahapan diantaranya normalisasi data latih, perhitungan jarak antara data latih terhadap data uji, pengurutan jarak terkecil hingga terbesar, pengambilan kelas sebanyak K dan terakhir penentuan kelas.

Pengujian menggunakan dataset sebanyak 200 yang terdiri dari 180 data latih dan 20 data uji. Ujicoba menggunakan parameter nilai $K=3$, $K=5$ dan $K=7$, didapatkan akurasi tertinggi sebesar 95% yang dihasilkan ketika K bernilai 3.

DAFTAR PUSTAKA

- UPT Puskesmas Sumberasih. (2023). Profil Kesehatan UPT Puskesmas Sumberasih Tahun 2022. UPT Puskesmas Sumberasih.
- Penyakit Tidak Menular Indonesia. (2017, September 26). Carta Prediksi Resiko. Youtube. <https://www.youtube.com/watch?v=I6bZ9aNxfz0>
- Age Risna Susilo, 2018, Klasifikasi Penentuan Penerimaan Beras Miskin Menggunakan Decision Tree. Universitas Dian Nuswantoro, Semarang.
- Dwi Fasnuari, H. A., Yuana, H., & Chulkamdi, M. T. (2022). Penerapan Algoritma K-

- Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus. *Antivirus : Jurnal Ilmiah Teknik Informatika*, 16(2), 133–142.
<https://doi.org/10.35457/antivirus.v16i2.2445>
- Harun, R., Pelangi, K.C., & Lasena, Y. (2020). Penerapan Data Mining untuk Mentukan Potensi Hujan Harian Dengan Menggunakan Algoritma K-Nearest Neighbour (KNN). *Jurnal Manajemen Informatika dan Sistem Informasi*, 3(1), 8-15.
<https://doi.org/10.36595/MISI.V3I1.125>.
- Hasran. (2020). Klasifikasi Penyakit Jantung Menggunakan Metode K-Nearest Neighbor. *Indonesia Journal of Data and Science*, 1(1), 6–10.
<http://bit.ly/datasetcardio>.
- Hermawati, F. A. (2013). *Data Mining*. Andi Offset.
- Indriasih, E., Rosita, T., Yulianti, A., & Agustiya, R. I. (2020). Penilaian Kualitas Data Penyebab Kematian di Indonesia Tahun 2014. *Buletin Penelitian Kesehatan*, 48(4), 235–242. <https://doi.org/10.22435/bpk.v48i4.3524>
- Prasetyo, E. (2014). *Data Mining - Mengolah Data Menjadi Informasi*. ANDI.
- Ryfai, D. A., Hidayat, N., & Santoso, E. (2022). Klasifikasi Tingkat Resiko Serangan Penyakit Jantung menggunakan Metode K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(10), 4701–4707.
<http://j-ptiik.ub.ac.id>
- Santoso, B. R., Gaghauna, E. E. M., & Akbar, I. (2023). Prediksi Kejadian Penyakit Jantung Dan Pembuluh Darah Di Upt Puskesmas Rawat Inap Alabio. *Jurnal Persatuan Perawat Nasional Indonesia (JPPNI)*, 8(1), 1.
<https://doi.org/10.32419/jppni.v8i1.360>
- Sinambela L.P. (2014). *Reformasi Pelayanan Publik*. PT. Bumi Aksara.
- Sumberasih, U. P. (2022). *Laporan Kinerja Tahunan*.
- Susanto, H., & Jamal, J. (2020). Identifikasi Spesies Ikan Berdasarkan Kontur Otolith Menggunakan Metode Otsu Dan Back Propagation Neural Network. *Joutica*, 5(2).
<https://doi.org/10.30736/jti.v5i2.486>
- Yanto, D., Susanto, H., Zulkifli, K., & Gupron, F. R. (2023). Penerapan Algoritma K-Means Dalam Menentukan Kualitas Satuan Pendidikan Berdasarkan Nilai Internal Dan Eksternal. *JEECOM Journal of Electrical Engineering and Computer*, 5(2), 319–328. <https://doi.org/10.33650/jeecom.v5i2.6946>