Jurnal Informatika dan Rekayasa Perangkat Lunak ISSN 2656-2855
Vol. 7, No. 2, September 2025, Hal. 213-220 e-ISSN 2685-5518

Implementasi Algoritma K-Nearest Neighbor dalam Prediksi Penyakit Jantung

Arif Ardiansyah[1], **Juan Stanley Nou Tuandali**[2], **Latiful Sirri**[3], **Rinci Kembang Hapsari**[4*], **Syahrul Riza Andi Santoso** [5]
[1,2,3,4,5] Program Studi Teknik Informatika, Fakultas Teknik Elektro dan Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya
[*]Email: rincikembang@itats.ac.id

Abstrak

Penyakit gagal jantung merupakan masalah kesehatan yang serius dan mendesak yang mempengaruhi jutaan orang di seluruh dunia. Ada beberapa faktor yang mempengaruhi terjadinya gagal jantung, seperti usia, jenis nyeri data, tekanan darah, kadar kolesterol, dan faktor-faktor risiko lainnya yang berhubungan dengan penyakit jantung. Dengan perkembangan teknologi saat ini, data mining dan pembelajaran mesin dapat digunakan untuk memprediksi kondisi kesehatan pasien. Sehingga permasalahan penelitian ini adalah bagaimana mengimplementasikan teknik data maining untuk identifikasi penyakit jantung. Dan tujuan dari penelitian adalah mengidentifikasi penyakit jantung untuk mencegah terjadinya gagal jantung. Penelitian ini memanfaatkan algoritma K-Nearest Neighbor (k-NN) untuk memperkirakan kemungkinan pasien mengalami gagal jantung berdasarkan fitur data yang tersedia. Data yang digunakan diambil dari situs kaggle.com yang mencakup informasi dari pasien yang didiagnosis menderita penyakit gagal jantung dan mereka yang tidak menderita gagal jantung. Proses analisis melibatkan langkah-langkah pengolahan data, seperti normalisasi, pengelompokan fitur, dan pemilihan parameter K yang optimal untuk algoritma k-NN. Evaluasi yang dilakukan dengan menghitung nilai akurasi, presisi, recall, dan F1-score. Pengujian dilakukan pada dataset dengan 299 data pasien, yang dibagi menjadi data latih dan data uji dengan perbandingan 80:20. Hasil dari penelitian ini menunjukkan bahwa algoritma k-NN memiliki akurasi sebesar 87% dalam memprediksi penyakit gagal ginjal. Hasil ini menunjukkan bahwa algoritma k-Nearest Neighbor dapat memprediksi penyakit gagal jantung dengan baik

***Kata kunci**: k-Nearest Neighbor, gagal ginjal, akurasi, prediksi*

Abstract

Heart failure is a serious and pressing health problem that affects millions of people worldwide. Several factors influence the occurrence of heart failure, such as age, type of pain, blood pressure, cholesterol levels, and other risk factors associated with heart disease. With current technological developments, data mining and machine learning can be used to predict patient health conditions. Therefore, the problem of this research is how to implement data mining techniques for identifying heart disease. The goal of the study is to identify heart disease and prevent heart failure. This study utilises the K-Nearest Neighbour (k-NN) algorithm to estimate the likelihood of patients experiencing heart failure based on available data features. The data used is taken from the kaggle.com site, which includes information from patients diagnosed with heart failure and those who do not suffer from heart failure. The analysis process involves data processing steps, such as normalisation, feature grouping, and selecting the optimal K parameter for the k-NN algorithm. Evaluation is carried out by calculating the accuracy, precision, recall, and F1-score values. Testing is carried out on a dataset with 299 patient data, which is divided into training data and test data with a ratio of 80:20. The results of this study indicate that the k-NN algorithm has an accuracy of 87% in predicting kidney failure. This result indicates that the k-Nearest Neighbour algorithm can effectively predict heart failure.

Keywords: k-Nearest Neighbor, kidney failure, accuracy, predict

PENDAHULUAN

Gagal jantung merupakan kelainan multisistem dimana terjadi gangguan pada jantung, otot skelet dan fungsi ginjal, stimulasi sistem saraf simpatis serta aktivitas hormonal yang dapat mengancam jiwa. Penyakit ini disebabkan oleh banyak faktor risiko seperti hipertensi, diabetes melitus, dislipidemia, obesitas, dan gaya hidup yang tidak sehat (Damara and Ariwibowo, 2021). Selain itu, berdasarkan data dari Kementerian Kesehatan Republik Indonesia, tercatat bahwa pada tahun 2015 terdapat 70% kematian di dunia disebabkan oleh Penyakit Tidak Menular, dimana 45% di antaranya disebabkan oleh penyakit jantung dan pembuluh darah, yaitu 17.7 juta dari 39,5 juta kematian.

Identifikasi dini dan pencegahan faktor risiko penyakit jantung koroner sangat penting untuk menurunkan angka kesakitan dan kematian akibat penyakit ini. Dalam menghadapi kompleksitas penyakit kardio-vaskular, pendekatan inovatif diperlukan. Saat ini, kemajuan di bidang teknologi informasi dan pembelajaran mesin telah memungkinkan pengembangan model prediksi penyakit jantung yang lebih akurat dan efisien. Data mining merupakan proses penggalian informasi yang berguna dari sekumpulan data besar (Nalatissifa et al., 2021) (Zuriati & Qomariyah, 2022). Dalam konteks medis, data mining dapat dimanfaatkan untuk mengana-lisis data pasien dan mengidentifikasi pola serta tren yang dapat membantu dalam membuat keputusan perawatan kesehatan yang lebih baik (Nalatissifa et al., 2021). Salah satu cabang dari data mining yang sering digunakan adalah prediksi atau klasifikasi (Nalatissifa et al., 2021) (Azis et al., 2020) (Silvana et al., 2020).

Teknik klasifikasi merupakan teknik data mining yang bertujuan untuk memprediksi kelas atau kategori suatu objek berdasarkan karakteristik atau atribut yang dimiliki. Klasifikasi merupakan teknik supervise learning, di mana model pembelajaran dilatih menggunakan dataset yang sudah dilabeli dengan kelas tertentu. Selain itu, pemanfaatan teknologi data mining juga dapat membantu dalam mengklasifikasikan jenis penyakit kardiovaskular yang diderita oleh pasien berdasarkan fitur-fitur yang dimiliki, seperti karakteristik demografi, gaya hidup, dan riwayat kesehatan (Dewi et al., 2022).

Studi sebelumnya menunjukkan bahwa pengembangan model prediksi penyakit jantung dengan menggunakan data medis pasien telah dilakukan dengan beberapa algoritma machine learning. Penelitian prediksi penyakit jantung menggunakan algoritma *Decision Tree*, dan *Random Forest* yang telah dilakukan sebelumnya telah mencapai tingkat akurasi yang tinggi, dengan tingkat keberhasilan di atas 80% (Sugriyono & Siregar, 2020)(Erlin et al., 2022)(Megawaty & Huda, 2021). Prediksi penyakit jantung juga telah digunakan dengan algoritma *Logistic Regression* dengan menghasilkan nilai akurasi yang baik (Erlin et al., 2022). Penelitian lain juga telah menggunakan algoritma *Support Vector Machine*, dan *Neural Network* dengan tingkat akurasi yang bervariasi (Damara & Ariwibowo, 2021)(Namli, 2021)(Devi et al., 2023)(Girianto, 2020).

Dalam penelitian lain, algoritma k-Nearest Neighbor telah diterapkan untuk memprediksi penyakit stroke dan menghasilkan tingkat akurasi yang tinggi (Zuriati & Qomariyah, 2022). Algoritma k-NN juga digunakan untuk memprediksi penyakit diabetes pada penelitian sebelumnya memberikan tingkat akurasi yang sangat tinggi, mencapai 89%.(Anggrawan & Mayadi, 2023). Beberapa penelitian sebelumnya juga menunjukkan bahwa algoritma k-NN dapat digunakan untuk mengidentifikasi dan mengklasifikasikan pasien kanker dan penyakit lainnya dengan baik (Nalatissifa et al., 2021)(Naufal et al., 2020), dan untuk mendeteksi penyakit ginjal kronis(Wijaya *et al.*, 2024). K-NN merupakan algoritma pembelajaran mesin yang sederhana namun efektif, yang mengklasifikasikan kasus baru berdasarkan fitur dari k sampel terdekat dalam dataset.

Algoritma k-NN juga telah banyak digunakan untuk melakukan prediksi data kesehatan dan menunjukkan kinerja yang menjanjikan (Ankireddy, 2019) (Alfando & Hayami, 2023) (Erlin et al., 2022). Dalam penelitian ini, kami akan mengeksplorasi potensi penggunaan algoritma k-NN untuk memprediksi kemungkinan pasien menderita penyakit jantung berdasarkan karakteristik klinis mereka, seperti usia, jenis kelamin, tekanan darah, dan kolesterol. Dengan penelitian ini, diharapkan dapat memberikan wawasan berharga bagi praktisi kesehatan untuk mengidentifikasi pasien

berisiko tinggi dan memfasilitasi intervensi dini yang efektif.

TINJAUAN PUSTAKA

Penyakit jantung adalah gangguan yang memengaruhi fungsi jantung dan pembuluh darah, sering kali disebabkan oleh penyumbatan, kerusakan, atau gangguan pada sistem kardiovaskular. Penyakit ini merupakan salah satu penyebab utama kematian di dunia.

Menurut data dari Institute for Health Metrics and Evaluation (IHME) tahun 2019, Indonesia mencatat 651.481 kematian per tahun akibat penyakit kardiovaskular. Rinciannya meliputi: Stroke: 331.349 kematian, Penyakit jantung koroner: 245.343 kematian, Penyakit jantung hipertensi: 50.620 kematian, dan Penyakit kardiovaskular lainnya: sisanya

Berdasarkan Organisasi Kesehatan Dunia (WHO) melaporkan bahwa lebih dari 17 juta orang di seluruh dunia meninggal setiap tahunnya akibat penyakit kardiovaskular.

Data mining adalah proses pengumpulan dan pengolahan data dalam jumlah besar untuk mengekstraksi informasi penting yang tersembunyi di dalamnya. Proses ini melibatkan penggunaan teknik statistik, matematika, dan kecerdasan buatan untuk mengidentifikasi pola atau hubungan yang tidak langsung terlihat. Data mining sering disebut juga sebagai Knowledge Discovery in Databases (KDD), yang mencakup serangkaian tahapan mulai dari pembersihan data, integrasi, transformasi, hingga evaluasi pola yang ditemukan.

Tujuan utama dari data mining adalah untuk membantu pengambilan keputusan dengan menyediakan wawasan yang berharga dari data yang tersedia. Dalam dunia bisnis, misalnya, data mining dapat digunakan untuk memahami perilaku pelanggan, memprediksi tren pasar, atau mendeteksi penipuan. Metode yang umum digunakan dalam data mining meliputi asosiasi, klasifikasi, regresi, dan clustering, yang masing-masing memiliki peran spesifik dalam analisis data. Dengan penerapan yang tepat, data mining dapat menjadi alat yang sangat efektif dalam mengoptimalkan strategi dan operasional di berbagai sektor

Pre-processing data adalah tahap awal dalam proses data mining yang bertujuan untuk mempersiapkan data mentah menjadi format yang siap untuk dianalisis. Langkah ini mencakup pembersihan data, integrasi, transformasi, dan reduksi data. Pembersihan data melibatkan penanganan nilai yang hilang, duplikasi, dan inkonsistensi. Integrasi data menggabungkan data dari berbagai sumber menjadi satu kesatuan yang konsisten. Transformasi data mencakup normalisasi dan pengkodean ulang,

Implementasi pre-processing yang efektif sangat penting karena kualitas data yang baik secara langsung mempengaruhi akurasi dan efektivitas model analitik yang dibangun. Dengan data yang telah diproses dengan baik, proses penambangan data dapat menghasilkan wawasan yang lebih valid dan dapat diandalkan.

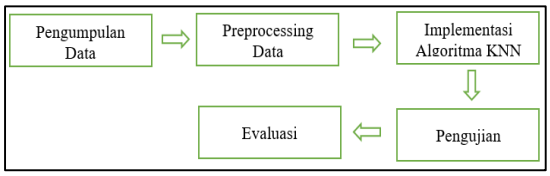
Klasifikasi dalam data mining adalah proses pengelompokan data ke dalam kategori atau kelas yang telah ditentukan berdasarkan karakteristik tertentu. Ini adalah metode umum yang digunakan untuk memperkirakan kelas dari suatu objek yang labelnya belum diketahui. Klasifikasi termasuk dalam kategori supervised learning, di mana model dibangun dari data pelatihan yang telah diberi label untuk mengklasifikasikan data baru.

Tujuan utamanya adalah memprediksi kelas atau kategori dari data baru yang belum diketahui labelnya. Proses ini melibatkan dua tahap utama: pertama, fase pembelajaran (*training*) di mana model dibangun menggunakan dataset yang sudah dilabeli; kedua, fase pengujian (*testing*) di mana model yang telah dibangun digunakan untuk mengklasifikasikan data baru. Klasifikasi sering diterapkan dalam berbagai bidang, seperti segmentasi pelanggan, deteksi penipuan, dan diagnosis medis.

Beberapa metode yang umum digunakan dalam klasifikasi meliputi: Decision Tree Analysis: Membangun model berupa pohon keputusan yang memetakan observasi tentang data ke kesimpulan target. Neural Networks: Menggunakan jaringan saraf tiruan yang meniru cara kerja otak manusia untuk mengenali pola kompleks dalam data. Support Vector Machines (SVM): Mencari hyperplane optimal yang memisahkan kelas-kelas dalam data dengan margin terbesar. Naïve Bayes: Menggunakan teorema Bayes dengan asumsi independensi antar fitur untuk mengklasifikasikan data. Dan k-Nearest Neighbors (k-NN): Mengklasifikasikan data baru berdasarkan kedekatannya dengan sejumlah k tetangga terdekat dalam dataset.

METODE

Kerangka penelitian mencangkup langkah-langkah yang dilakukan dalam pelaksanaan penelitian, ditunjukkan pada Gambar 1.



Gambar 1. Alur penelitian

3.1 Pengumpulan Data

Proses pengumpulan data awal kami dapatkan dari situ kaggle terkait dengan alamat: https://www.kaggle.com/datasets/ andrewmvd/heart-failure-clinical-data/data.

Data tersebut berisi 10 parameter seperti: usia (*age*), anemia, *creatine phosphokinase*, diabetes, *ejection fraction*, *blood pressure*, *platelets*, *serum creatine*, *serum sodium*, *sex*, *smoking*, dan *time*. Dan data label pada dataset tersebut adalah death event, yang menunjukkan kejadian gagal jantung dialami pansien sehingga menyebabkan kematian.

3.2 *Preprocessing* Data

Tahap pre-processing data merupakan tahap awal yang penting dalam proses data mining. Pada tahap pemrosesan data awal (preprocessing data), beberapa hal yang perlu dilakukan antara lain:

[1] Pembersihan Data (*Cleaning Data*) digunakan untuk menangani data yang tidak lengkap, tidak konsisten, atau mengandung kesalahan;
[2] Mengidentifikasi dan menangani data yang hilang (*missing values*).
[3] Mengubah format data agar sesuai untuk pemrosesan lebih lanjut
[4] Melakukan transformasi data seperti normalisasi atau standarisasi

Tahapan *pre-processing* ini akan memastikan data yang akan digunakan dalam proses data mining berkualitas baik dan siap untuk dianalisis lebih lanjut. Dalam penelitian ini dilakukan *cleaning* data, penanganan data hilang, serta transformasi data untuk mempersiapkan dataset yang akan digunakan untuk pelatihan dan pengujian model algoritma k-NN.

Dalam penelitian ini dilakukan seleksi fitur untuk mengurangi dimensi data dan memilih fitur-fitur yang paling berpengaruh dalam mengklasifikasikan penyakit jantung. Algoritma k-NN sangat sensitif terhadap dimensi data, sehingga seleksi fitur yang tepat dapat meningkatkan akurasi prediksi.

Transformasi data dilakukan dengan melakukan normalisasi (Azis et al., 2020) fitur-fitur yang digunakan agar berada dalam rentang nilai yang sama. Metode normalisasi(Siswa, 2023)(Hapsari et al., 2023) yang digunakan adalah metode *min-max value*, agar nilai parameter hanya berkisar antara 0 sampai 1, dimana perhitungan menggunakan persamaan (1).

$$x' = \frac{x - min(x)}{max(x) - min(x)} \quad (1)$$

Dimana x' = nilai baru, x = nilai x lama, min (x) = nilai paling kecil parameter x, max (x) = nilai paling besar pada parameter x.Sedangkan untuk seleksi fitur, kami menggunakan metode *Pearson Correlation* untuk menentukan paramater mana yang mempunyai korelasi positif dengan variabel target (Zuriati & Qomariyah, 2022) (Azis et al., 2020).

3.3 Algoritma k- Nearest Neighbor

K-Nearest Neighbor (k-NN) merupakan salah satu algoritma klasifikasi yang populer dalam data mining (Winantu & Khatimah, 2023). k-NN merupakan algoritma pembelajaran mesin yang sederhana namun efektif untuk klasifikasi dan regresi. Algoritma ini mengelompokan suatu data baru berdasarkan jarak data baru tersebut dengan k buah data tetangga terdekatnya dalam dataset (Zuriati & Qomariyah, 2022). Algoritma k-NN bekerja dengan mencari k buah data terdekat dari data baru, kemudian mengklasifikasikannya berdasarkan kelas mayoritas dari k buah data tersebut (Winantu & Khatimah, 2023).

Algoritma k-NN memiliki beberapa keunggulan, diantaranya:

1) Sederhana dan mudah diimplementasikan.
2) Dapat diterapkan pada berbagai macam kasus, baik klasifikasi maupun regresi.
3) Tidak memerlukan asumsi mengenai distribusi data.
4) Dapat menangani data yang berdimensi tinggi.
5) Memiliki kemampuan yang baik dalam menangani data yang tidak lengkap.

Selain keunggulan, algoritma k-NN juga memiliki beberapa kelemahan, yaitu:
1) Performa algoritma sangat bergantung pada pemilihan nilai k yang tepat.
2) Rentan terhadap dimensionalitas data yang tinggi (*curse of dimensionality*).
3) Membutuhkan komputasi yang intensif saat data training besar.

Meskipun demikian, algoritma k-NN tetap populer dan banyak digunakan dalam berbagai aplikasi, termasuk dalam bidang medis, karena kesederhanaan dan efektivitasnya yang baik. Langkah-langkah Algoritma k-NN, yaitu :
1) Menentukan nilai k
2) Menghitung jarak antara data baru dengan semua data latih menggunakan jarak Euclidean. Perhitungan jarak menggunakan persamaan (2).

$$dis = \sqrt{\sum_{i=0}^{n}(x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2 + \cdots} \quad (2)$$

3) Mengurutkan jarak-jarak tersebut dari jarak terpendek ke jarak terjauh
4) Menentukan k tetangga terdekat
5) Mengklasifikasikan data baru berdasarkan kelas mayoritas dari k tetangga terdekat
6) Mengevaluasi hasil klasifikasi (Zuriati & Qomariyah, 2022), (Azis et al., 2020), (Winantu & Khatimah, 2023), (Karepesina & Zahrotun, 2023).

3.4 Pengujian

Pada tahap pengujian, data dibagi menjadi 2 bagian, yaitu data latih dan data uji. Perbandingan data latih dan data uji adalah 80% dan 20%. Data uji dan data latih disediakan menggunakan metode *split validation.*

3.5 Evaluasi

Pada tahap ini dilakukan evaluasi terhadap hasil prediksi gagal jantung dengan algoritma k-NN. Jika hasilnya kurang dari 80% maka kami mulai lagi untuk melakukan perbaikan pada pre-processing data, seleksi fitur sampai perubahan nilai K

HASIL DAN PEMBAHASAN

Jumlah data pada dataset yang digunakan adalah 299 data. Dataset tersebut terdiri dari 12 parameter dengan label Death_Event sebagai label klasifikasi. Pada semua record, terlebih dahulu kami lakukan cleaning, untuk memastikan bahwa tidak ada data yang tidak lengkap. Dari proses tersebut kami dapati bahwa 100% data adalah lengkap, ditunjukkan pada Tabel 1.

Terlihat pada Table 1, atribut anemia, diabetes, *sex*, *smoking*, bernilai 0 dan 1. Sedangkan nilai atribut yang lain memiliki rentang yang sangat beragam dan lebar. Agar tidak terjadi bias, dilakukan proses normalisasi data dengan melakukan scaling pada parameter yang rentang nilainya beragam, sedemikian hingga nilainya menjadi antara 0 – 1. Normalisasi yang dilakukan menggunakan metode min-max scaler.

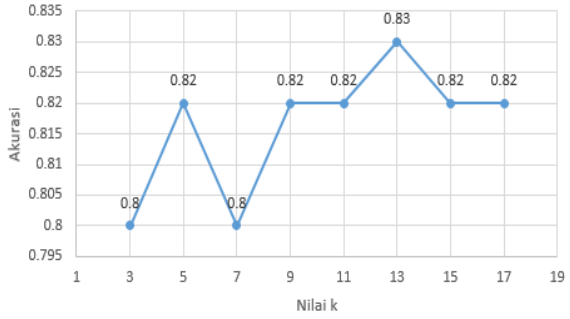
Dataset yang digunakan dalam penelitian ini dibagi menjadi dua bagian, yaitu sebagai data latih dan sebagai data uji, dengan persentase 80% dibanding 20%. Dengan jumlah data adalah 299, maka data latih sejumlah 239 data, dan data uji sejumlah 60 data.

Pada proses pengujian kami menggunakan nilai k adalah bilangan ganjil antara 3 sampai 17. Nilai bilangan ganjil diperlukan agar pada saat menentukan tetangga yang paling dekat, tidak terjadi kebingugan dalam menentukan kelas. Berdasarkan dataset dan scenario uji yang digunakan didapat nilai akurasi untuk setiap scenario ditunjukkan pada Gambar 2.

Tabel 1. Dataset Pasien Gagal Ginjal

No	age	anaemia	cp	Dia-betes	ef	Hbp	Plate-lets	sc	ss	sex	Smok-ing	time	de
1	75	0	582	0	20	1	265000	1.9	130	1	0	4	1
2	55	0	7861	0	38	0	263358	1.1	136	1	0	6	1
3	65	0	146	0	20	0	162000	1.3	129	1	1	7	1
4	50	1	111	0	20	0	210000	1.9	137	1	0	7	1
5	65	1	160	1	20	0	327000	2.7	116	0	0	8	1
.	….	….	….	….	…	…	…	…	…	…	….	…	…
299	75	1	246	0	15	0	127000	1.2	137	1	0	10	1

cp = creatinine phosphokinase, ef = ejection fraction, hbp = high blood pressure,
sc = serum creatinine ss = serum sodium, de = death event



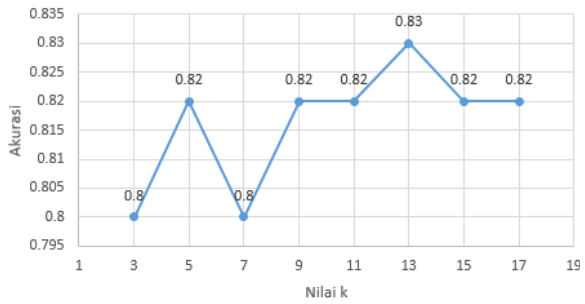
Gambar 2. Akurasi pengujian dengan 12 parameter

Pada Gambar 2, skenario pengujian dilakukan dengan 12 parameter, dan nilai akurasi tertinggi pada saat nilai k adalah 13. Pada skenrio pengujian yang kedua, menggunakan menggunakan seleksi fitur, dimana memilih parameter – parameter yang mempunyai korelasi tertinggi terhadap variabel target.

```
age                  0.253729
ejection_fraction    0.268603
serum_creatinine     0.294278
time                 0.526964
DEATH_EVENT          1.000000
Name: DEATH_EVENT, dtype: float64
```

Gambar 3. Nilai parameter dengan korelasi paling tinggi terhadap variabel target

Pada proses seleksi fitur, menggu-nakan algoritma *Pearson Correlation*, dimana bisa menentukan korelasi antara parameter dengan variabel target. Adapun hasil dari metode *Pearson Correlation* untuk setiap parameter dengan nilai diatas 0.2 ditunjukkan pada Gambar 3.

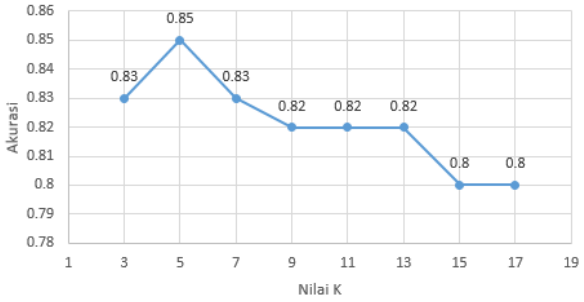


Gambar 4. Akurasi pengujian dengan 4 parameter hasil seleksi fitur

Sehingga, dari data korelasi diatas, terdapat 4 parameter yang akan kita gunakan pada pengujian selanjutnya, yaitu *age, ejection fraction, serum creatinine, serum sodium* dan *time*.

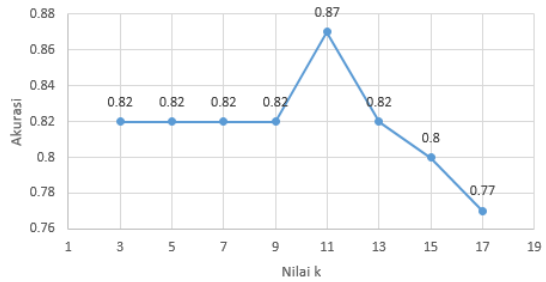
Dari Gambar 4 dengan empat parameter yang punya korelasi tertinggi, nilai akurasi untuk siap k hasil sama dengan sebelum dilakukan seleksi fitur.

Pada skenario pengujian ketiga, menggunakan tiga parameter tertinggi yaitu: *ejection fraction, serum creatinine* dan *time*. Nilai akurasi untuk setiap k ditunjukkan pada Gambar 5.



Gambar 5. Akurasi pengujian dengan tiga parameter hasil seleksi fitur

Dari pengujian terakhir didapatkan bahwa dengan 3 parameter yang punya korelasi paling tinggi, terjadi peningkatan akurasi sebesar 2 persen dari akurasi sebelumnya. Nilai k optimal juga berubah, dari yang sebelumnya k = 13, menjadi k = 5.



Gambar 6. Akurasi pengujian dengan tiga parameter hasil seleksi fitur, dengan dataset yang seimbang

Berdasarkan keseimbangan dataset terhadap jumlah data kejadian gagal ginjal yang terjadi Death event =1 dibandingkan dengan gagal jantung tidak terjadi kematian (Death event =0). Dimana Death event = 1, berjumlah 96 data, sedangkan data Death event = 0, sejumlah 203 data, sehingga data tidak seimbang.

Pada scenario pengujian ke-empat, dilakukan balancing data dengan melakukan eliminasi pada data yang nilai Death event = 0, sehingga jumlah data menjadi 96 data. Sehingga jumlah data dengan Death event = 1 sama dengan jumlah data dengan Death event = 0. Hasil Akurasi pengujian ditunjukkan pada Gambar 6.

Terlihat pada Gambar 6 bahwa akurasi tertinggi adalah 0,87 dengan nilai k = 11. Terdapat kenaikan akurasi sebesar 2%, dibandingkan ketika dataset tidak seimbang.

SIMPULAN

Berdasarkan hasil pengujian dapat disimpulkan bahwa pre-processing data memberikan peningkatan akurasi algoritma. Pre-Processing data yang dilakukan dalam penelitian ini adalah normalisasi, seleksi fitur dan penyeimbang data.

Proses seleksi fitur dengan menggunakan *Pearson Correlation* menghasilkan peningkatan akurasi sebesar 2% dari maksimal 83% menjadi 85%. Dan eliminasi terhadap data, agar data lebih berimbang, memberikan peningkatan akurasi 2% dari sebelumnya 85% menjadi 87%. Pada penelitian ini membuktikan bahwa algoritmat k-Nearest Neighbor mempunyai hasil yang cukup bagus dalam memprediksi terjadinya gagal jantung pada pasien.

Penelitian selanjutnya dapat dilakukan untuk meningkatkan performa algoritma K-NN, pada proses pengukuran *distance* bisa dibandingkan dengan *metric learning* (misalnya LMNN, *Mahalanobis learned*)

DAFTAR PUSTAKA

Alfando, A., & Hayami, R. (2023). KLASIFIKASI TEKS BERITA BERBAHASA INDONESIA MENGGU-NAKAN MACHINE LEARNING DAN DEEP LEARNING: STUDI LITERA-TUR. In JATI (Jurnal Mahasiswa Teknik Informatika) (Vol. 7, Issue 1, p. 681). https://doi.org/10.36040/jati.v7i1.6486

Anggrawan, A., & Mayadi, M. (2023). Application of KNN Machine Learning and Fuzzy C-Means to Diagnose Diabetes. In Matrik Jurnal Manajemen Teknik Informatika dan Rekayasa Komputer (Vol. 22, Issue 2, p. 405). https://doi.org/10.30812/matrik.v22i2.2777

Ankireddy, S. (2019). A Novel Approach to the Diagnosis of Heart Disease using Machine Learning and Deep Neural Networks. https://doi.org/10.1109/urtc49097.2019.9660581

Azis, H., Purnawansyah, P., Fattah, F., & Putri, I. P. (2020). Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung. In ILKOM Jurnal Ilmiah (Vol. 12, Issue 2, p. 81). https://doi.org/10.33096/ilkom.v12i2.507.81-86

Damara, C., & Ariwibowo, D. D. (2021). Diabetes Melitus tipe 2 sebagai faktor risiko penyakit jantung koroner (PJK) di RSUD Raden Mattaher Jambi tahun 2019. In Tarumanagara Medical Journal (Vol. 3, Issue 2, p. 249). https://doi.org/10.24912/tmj.v4i1.13715

Devi, N. L. P. L., Setiabudi, I. K., Harditya, K. B., & Wicaksana, I. G. A. T. (2023). Pelatihan tentang Resusitasi Jantung Paru (RJP) untuk Siswa SMA Guna Membentuk Remaja Tanggap Henti Jantung. In Jurnal Abdimas Kesehatan (JAK) (Vol. 5, Issue 2, p. 287). https://doi.org/10.36565/jak.v5i2.507

Dewi, P., Purwono, P., & Dwi, S. K. (2022). Pemanfaatan Teknologi Machine Learning pada Klasifikasi Jenis Hipertensi Berdasarkan Fitur Pribadi. In Smart Comp Jurnalnya Orang Pintar Komputer (Vol. 11, Issue 3). Politeknik Harapan Bersama Tegal. https://doi.org/10.30591/smartcomp.v11i3.3721

Erlin, E., Marlim, Y. N., Junadhi, Suryati, L., & Agustina, N. (2022). Deteksi Dini Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression. In Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI) (Vol. 11, Issue 2, p. 88). Gadjah Mada University. https://doi.org/10.22146/jnteti.v11i2.3586

Girianto, P. W. R. (2020). Pemberian Feedback pada Home Learning CPR untuk Meningkatkan Kemampuan Bystander CPR. In Jurnal Ners dan Kebidanan (Journal of Ners and Midwifery) (Vol. 7, Issue 1, p. 30).

https://doi.org/10.26699/jnk.v7i1.art.p030-036

Hapsari, R. K., Salim, A. H., Meilani, B. D., Indriyani, T., & Rachman, A. (2023). Comparison of the Normalization Method of Data in Classifying Brain Tumors with the k-NN Algorithm. In Advances in intelligent systems research/Advances in Intelligent Systems Research (p. 21). Atlantis Press. https://doi.org/10.2991/978-94-6463-174-6_3

Karepesina, F., & Zahrotun, L. (2023). Penerapan Data Mining Untuk Penentuan Penerima Beasiswa Dengan Metode K-Nearest Neighbor (K-NN). In Techno (Jurnal Fakultas Teknik Universitas Muhammadiyah Purwokerto) (Vol. 24, Issue 1, p. 1). Muhammadiyah University Purwokerto. https://doi.org/10.30595/techno.v24i1.9084

Megawaty, M., & Huda, N. (2021). Pembaharuan Sistem Penentuan Untuk Klasifikasi Jenis Penyakit pada RSUD Sekayu Menggunakan Pendekatan Extreme Programming. In JURNAL MEDIA INFORMATIKA BUDIDARMA (Vol. 5, Issue 1, p. 66). https://doi.org/10.30865/mib.v5i1.2273

Nalatissifa, H., Gata, W., Diantika, S., & Nisa, K. (2021). Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja. In Jurnal Informatika Universitas Pamulang (Vol. 5, Issue 4, p. 578). https://doi.org/10.32493/informatika.v5i4.7575

Namli, S. (2021). HUBUNGAN KONSEN-TRASI HEMOGLOBIN DARAH DENGAN KEJADIAN INFARK MIO-KARD AKUT DI RUMAH SAKIT UNIVERSITAS SUMATERA UTARA PERIODE 2018-2019. In JIMKI Jurnal Ilmiah Mahasiswa Kedokteran Indonesia (Vol. 9, Issue 2, p. 20). https://doi.org/10.53366/jimki.v9i2.468

Naufal, S. A., Adiwijaya, A., & Astuti, W. (2020). Analisis Perbandingan Klasifikasi Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN) untuk Deteksi Kanker dengan Data Microarray. In JURIKOM (Jurnal Riset Komputer) (Vol. 7, Issue 1, p. 162). https://doi.org/10.30865/jurikom.v7i1.2014

Silvana, M., Akbar, R., & Alfi, S. (2020). Pemanfaatan Metode Naïve Bayes dalam Implementasi Sistem Pakar Untuk Menganalisis Gangguan Perkembangan Anak. In Jurnal Nasional Teknologi dan Sistem Informasi (Vol. 6, Issue 2, p. 74). Andalas University. https://doi.org/10.25077/teknosi.v6i2.2020.74-81

Siswa, T. A. Y. (2023). Komparasi Optimasi Chi-Square, CFS, Information Gain dan ANOVA dalam Evaluasi Peningkatan Akurasi Algoritma Klasifikasi Data Performa Akademik Mahasiswa. In Informatika Mulawarman Jurnal Ilmiah Ilmu Komputer (Vol. 18, Issue 1, p. 62). https://doi.org/10.30872/jim.v18i1.11330

Sugriyono, S., & Siregar, M. U. (2020). Preprocessing kNN algorithm classification using K-means and distance matrix with students' academic performance dataset. In Jurnal Teknologi dan Sistem Komputer (Vol. 8, Issue 4). Diponegoro University. https://doi.org/10.14710/jtsiskom.2020.13874

Wijaya, A.K. et al. (2024) 'Identifikasi Penyakit Ginjal Kronis Menggunakan Algoritma K-Nearest Neighbour (k-NN)', in Seminar Nasional Informatika Bela Negara (SANTIKA) ISSN, pp. 361–365.

Winantu, A., & Khatimah, C. (2023). Perbandingan Metode Klasifikasi Naive Bayes Dan K-Nearest Neighbor Dalam Memprediksi Prestasi Siswa. In INTEK Jurnal Informatika dan Teknologi Informasi (Vol. 6, Issue 1, p. 58). https://doi.org/10.37729/intek.v6i1.3006

Zuriati, Z., & Qomariyah, N. (2022). Klasifikasi Penyakit Stroke Menggunakan Algoritma K-Nearest Neighbor (KNN). In ROUTERS Jurnal Sistem dan Teknologi Informasi (p. 1). https://doi.org/10.25181/rt.v1i1.2665