

Komparasi Penerapan Algoritma C4.5, K-Nearest Neighbor, dan Naïve Bayes untuk Keberlangsungan Pasien Gagal Jantung

Muhammad Fakhri Rizqullah, Naura Tri Raihana, Muhammad Ihsan Jambak*

Fakultas Ilmu Komputer, Sistem Informasi, Universitas Sriwijaya, Palembang, Indonesia

Email: ¹09031382126124@student.unsri.ac.id, ²09031282126063@student.unsri.ac.id, ^{3,*}jambak@unsri.ac.id

Email Penulis Korespondensi: jambak@unsri.ac.id

Abstrak—Angka total kematian dunia diakibatkan gagal jantung terus menunjukkan peningkatan. Klasifikasi pasien dengan akurasi terbaik dapat membantu meningkatkan langkah pencegahan berdasarkan informasi klinis. Penelitian ini melakukan komparasi algoritma klasifikasi antara C4.5, K-Nearest Neighbor, dan Naïve Bayes didasarkan pada CRISP-DM dengan teknik evaluasi model *10-fold cross validation* dan *pairwise t-test* menggunakan *software* RapidMiner. Hasil penelitian mendapatkan nilai akurasi tertinggi dengan nilai 0,779 dengan standar deviasi kurang lebih 0,046. Hasil penelitian menunjukkan bahwa algoritma C4.5 memiliki performa terbaik, disusul oleh algoritma Naïve Bayes dengan perbedaan yang tidak cukup signifikan dan terakhir algoritma K-Nearest Neighbor dengan nilai terkecil sehingga dinilai kurang baik diimplementasikan dalam dataset.

Kata Kunci: Gagal jantung; Algoritma; Klasifikasi; Komparasi; C4.5

Abstract—The total number of deaths worldwide due to heart failure continues to show an increase. Classifying patients with the best accuracy can help improve preventive measures based on clinical information. This study compares classification algorithms including C4.5, K-Nearest Neighbor, and Naïve Bayes based on CRISP-DM with the 10-fold cross-validation model evaluation technique and pairwise t-test using RapidMiner software. The research obtained the highest accuracy value of 0.779 with a standard deviation of approximately 0.046. The research results indicate that the C4.5 algorithm performs the best, followed by the Naïve Bayes algorithm with a statistically insignificant difference, and lastly, the K-Nearest Neighbor algorithm with the smallest value, thus considered less suitable for implementation in the dataset.

Keywords: Heart failure; Algorithm; Classification; Comparison; C4.5

1. PENDAHULUAN

Seluruh makhluk hidup pasti mendambakan kesehatan dan umur panjang. Kesehatan menjadi faktor utama dalam menentukan tujuan dan prestasi setiap individu dalam hidup. Berusaha menjadi produktif secara jasmani merupakan upaya bagi kita agar terhindar dari berbagai penyakit, salah satunya penyakit kardiovaskular [1]. Penyakit kardiovaskular adalah jenis penyakit yang mencakup masalah pada jantung dan pembuluh darah. Statistik dunia mengatakan, ada 9,4 juta jiwa manusia meninggal akibat penyakit kardiovaskular. Sebagai salah satu organ vital manusia yang bertugas memompa darah, sebagian orang yang didiagnosis memiliki penyakit kardiovaskular akan mengalami ketidaknyamanan sepanjang hidupnya [2]. Penyakit kardiovaskular paling umum dan sering terjadi merupakan gagal jantung. Gagal jantung merupakan suatu kegagalan otot jantung untuk memompakan darah secara memadai ke seluruh tubuh [3]. Berdasarkan data dari WHO, sekitar 17,5 juta orang meninggal dunia karena penyakit kardiovaskular, yang menyumbang sekitar 31% dari total kematian di seluruh dunia. [4]. Di Indonesia sendiri, mengutip dari pernyataan Riset Kesehatan Dasar (Riskedas) (2018) menyatakan bahwa prevalensi penyakit gagal jantung berada di persentase 1,5% dari total populasi atau sekitar 4.107.00 orang. Banyak faktor yang dapat menyebabkan penyakit gagal jantung seperti diabetes, hipertensi, pola makan tidak sehat, begadang, dan kurangnya aktifitas fisik [5]. Metode analisis konvensional secara manual dinilai tidak optimal dalam proses diagnosis, melainkan dengan bantuan pengetahuan terkomputerisasi dianggap lebih efektif dan relevan [6]. Klasifikasi dapat dilakukan dengan *data mining* yang merupakan proses pengelompokan informasi berdasarkan data yang besar dengan menggunakan prosedur analisis dan teknik statistik [5], [7], [8].

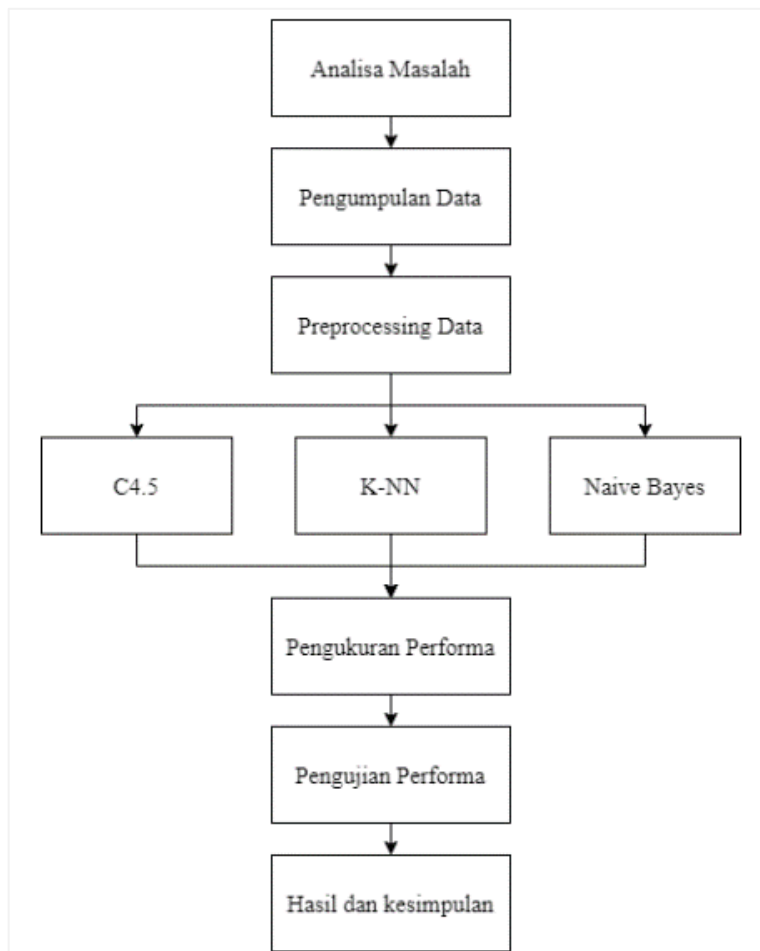
Dalam penelitian ini, peneliti melakukan klasifikasi dengan tujuan mengidentifikasi kelangsungan hidup pasien dengan diagnosis gagal jantung. Pada tahun 2021, penelitian dilakukan oleh Cahya untuk memprediksi gagal jantung menggunakan metode Naïve Bayes dengan hasil akurasi 96,67% [2]. Penelitian oleh Saifu, dkk melakukan klasifikasi K-Nearest Neighbor dengan kasus seleksi penerimaan beasiswa dengan hasil akurasi 90,5% [9]. Pada Tahun 2022, penelitian dilakukan oleh Sepharni, dkk untuk melakukan klasifikasi penyakit jantung menggunakan algoritma C4.5 dengan hasil akurasi 79% [10]. Penelitian oleh A'Yuninyah, dkk menerapkan implementasi Naïve Bayes Classifier (NBC) pada gangguan ginjal kronik dengan hasil akurasi 96,43% [11]. Penelitian oleh Subarkah, dkk pada tahun 2020 tentang prediksi nasabah bank dengan melakukan komparasi antara algoritma Decision Tree, K-Nearest Neighbor, dan Naïve Bayes mendapatkan hasil algoritma Decision Tree terbaik dengan akurasi 88,48%, Naïve Bayes dengan akurasi 86,86%, dan K-Nearest Neighbour dengan nilai paling kecil, yaitu 84,96% [12].

Berdasarkan penelitian sebelumnya, terdapat beberapa perbedaan hasil akurasi pada penerapan metode yang berbeda, maka dari itu untuk mengetahui algoritma yang memiliki prediksi paling tepat, peneliti akan melakukan komparasi dari metode *Decision Tree* Algoritma C4.5, K-Nearest Neighbors (K-NN), dan Naïve Bayes. Peneliti menggunakan tiga metode tersebut untuk mengevaluasi dan membandingkan metode mana yang lebih sesuai untuk diterapkan pada dataset dalam konteks klasifikasi kasus keberlangsungan pasien gagal jantung serta mendapatkan akurasi prediksi terbesar, mengingat jantung merupakan organ paling berharga maka diperlukan adanya model akurat yang dapat mengklasifikasikan gagal jantung berdasarkan informasi klinis untuk membantu tindakan pencegahan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Dalam menjalankan penelitian ini, peneliti mengikuti pendekatan CRISP-DM sebagai kerangka kerja penelitian. Berikut pada gambar 1 merupakan langkah-langkah prosedur yang dilakukan dalam penelitian:



Gambar 1. Metodologi Penelitian

Pada tahapan penelitian menjelaskan alur dari permasalahan yang telah dianalisa, lalu melakukan pengumpulan dan *preprocessing* data dan melakukan komparasi pada algoritma C4.5, K-NN, dan Naive Bayes. Setelah melakukan komparasi akan dilakukan pengukuran perbandingan performa dan pengujian untuk mengetahui hasil akhir tervalidasi. Proses pemahaman masalah dimulai dengan melakukan analisa mendalam. Berdasarkan permasalahan yang telah ditemukan oleh analisa diatas [13], maka penelitian ini dilakukan karena diperlukan adanya model akurat yang dapat mengklasifikasikan gagal jantung berdasarkan informasi klinis untuk membantu tindakan pencegahan.

2.2 Pengumpulan dan *Preprocessing* Data

Dalam konteks penelitian ini dan masalah yang dihadapi, diperlukan dataset yang sesuai untuk memastikan hasil yang akurat dengan algoritma yang digunakan. Dataset yang digunakan diambil dari situs UCI Machine Learning dengan judul Heart Failure Clinical Records Dataset yang dapat diakses melalui tautan <https://archive.ics.uci.edu/dataset/519/heart+failure+clinical+records>. Dataset ini terdiri dari rekam medis 299 pasien yang menderita gagal jantung, yang dikumpulkan selama masa follow-up. Setiap profil pasien mencakup 13 atribut klinis yang relevan dengan diagnosis gagal jantung. Dataset ini mencakup atribut yang merupakan faktor-faktor yang memengaruhi kelangsungan hidup pasien dengan gagal jantung.

Tabel 1. Informasi Dataset

Atribut	Keterangan
age	Umur pasien
anaemia	Penurunan hemoglobin (boolean)
creatinine_phosphokinase	Tingkatan enzim CPK pada darah (mcg/L)
diabetes	Pasien menderita diabetes

ejection_fraction	Persentase darah dalam kontaksi meninggalkan jantung
high_blood_pressure	Pasien menderita hipertensi
platelets	Trombosit darah (kiloplatelets/mL)
serum_creatinine	Tingkat kreatinin serum dalam darah (mg/dL)
serum_sodium	Kadar natrium serum dalam darah (mEq/L)
sex	Jenis kelamin laki-laki atau perempuan (binary)
smoking	Jika pasien merokok atau tidak (boolean)
time	Periode penanganan medis (hari)
death_event	Jika pasien meninggal dalam penanganan medis (boolean)

Pada tabel 1, dijelaskan bahwa setiap atribut memiliki keterangan tersendiri terkait informasi klinis yang akan dikumpulkan melalui pasien yang memiliki gejala pada penyakit jantung.

Tabel 2. Dataset Pasien

age	anaemia	creatinine	phospokinase	diabetes	ejection_fraction	high_blood_pressure	...	death_event
75	0	582		0	20	1	...	1
55	0	7861		0	38	0	...	1
65	0	146		0	20	1	...	1
50	1	111		0	20	1	...	1
...
50	0	196		0	45	1	...	0

Pada tabel 2, setiap atribut diisi dengan data klinis pasien yang telah disesuaikan agar data dapat diproses menggunakan algoritma dengan baik. Langkah selanjutnya adalah data *preparation* yang merupakan penyiapan data meliputi penyusunan, pengaturan, penggabungan, sehingga dapat digunakan dalam *machine learning* nantinya [14]. Data *preparation* merupakan langkah yang dilakukan agar data sesuai dengan model yang akan digunakan. Setelah dicek secara keseluruhan, tidak diperlukan adanya drop kolom, karena semua dataset memiliki pengaruh pada keberlangsungan pasien gagal jantung. Data juga tidak memiliki *missing value*, artinya tidak ada kolom yang NULL pada dataset ini. Hanya saja diperlukan transformasi data pada kolom anaemia, diabetes, high_blood_pressure, smoking, dan DEATH_EVENT memiliki nilai 1 dan 0. Hal ini tidak sesuai dengan perbandingan algoritma yang akan dipakai, maka diperlukan adanya transformasi data dengan nilai nol menjadi “No” dan satu menjadi “Yes”. Hal serupa juga diterapkan pada kolom sex dimana diperlukan transformasi data dengan nilai nol menjadi “Female” dan nilai satu menjadi “Male”. Hasil dari dataset setelah proses preparasi dapat dilihat pada tabel 3 dibawah:

Tabel 3. Dataset Pasien Setelah *Preparation*

age	anaemia	creatinine	phospokinase	diabetes	ejection_fraction	high_blood_pressure	...	death_event
75	No	582		No	20	Yes	...	Yes
55	No	7861		No	38	No	...	Yes
65	No	146		No	20	Yes	...	Yes
50	Yes	111		No	20	Yes	...	Yes
...
50	No	196		No	45	No	...	No

2.3 C4.5

Algoritma C4.5 merupakan implementasi dari teknik pohon keputusan yang digunakan dalam klasifikasi *data mining*. Secara umum, algoritma C4.5 membangun pohon keputusan dengan langkah-langkah berikut:

- Memilih atribut sebagai simpul akar.
- Membuat cabang untuk setiap nilai atribut.
- Memisahkan kasus ke dalam cabang berdasarkan nilai atribut.
- Mengulangi proses untuk setiap cabang hingga semua cabang memiliki kelas yang sama.

Nilai keuntungan dari setiap atribut menentukan simpul dalam pohon keputusan menggunakan suatu formula [15]:

$$Entropy(X) \sum_{i=1}^n - p_i * \log_2 p_i \tag{1}$$

Dengan keterangan X merupakan jumlah himpunan kasus, N merupakan jumlah partisi X, dan Pi merupakan proporsi dari Xi dan X.

$$Gain(X,A) = Entropy - \sum_{i=1}^n \frac{|X_i|}{|X|} * Entropy(X_i) \tag{2}$$

Dengan keterangan X merupakan jumlah himpunan kasus, A merupakan atribut, N merupakan jumlah partisi atribut A, |Xi| merupakan jumlah kasus partisi ke -I, dan |X| merupakan jumlah kasus dalam.

2.4 K-Nearest Neighbor (K-NN)

K-Nearest Neighbor atau yang sering disebut *lazy learning* adalah algoritma yang mengklasifikasikan *instance query* yang baru dengan mayoritas kategori yang dekat dalam K-NN itu sendiri. Nilai k mewakili jumlah data terdekat dari data uji, yang berarti k adalah jumlah data yang jaraknya paling dekat dengan objek [9]. Nilai k terbaik didasari oleh nilai sebuah data untuk mengurangi *noise* pada proses [16]. Hal ini juga dapat dilakukan menggunakan *Euclidean Distance* dengan persamaan:

$$d_{Euc}(x, y) = \sqrt{\sum_i^p (x1 - x2)^2} \quad (3)$$

Dengan keterangan d_{Euc} merupakan jarak ukur antara data sampel dan data uji, $x1$ merupakan sampel data, $x2$ merupakan data uji, P merupakan dimensi data, dan i merupakan variabel data.

2.5 Naïve Bayes (NB)

Naïve Bayes adalah algoritma klasifikasi yang memanfaatkan prinsip probabilitas untuk mengantisipasi kejadian di masa depan berdasarkan informasi atau pengalaman sebelumnya. Algoritma ini melakukan klasifikasi data dengan memperhitungkan probabilitas P atribut x dari setiap kelas y data, dengan asumsi bahwa semua atribut bersifat independen dan tidak saling terkait [17], dengan menggunakan persamaan yang diberikan:

$$P(H|X) = \frac{p(X|H)p(H)}{p(X)} \quad (4)$$

Dengan keterangan X merupakan jumlah data belum diketahui, H merupakan kelas spesifik hipotesis X , $P(H|X)$ merupakan probabilitas hipotesis H berdasarkan X , dan $P(H)$ merupakan probabilitas hipotesis H

2.6 Performance

Performance adalah parameter atau ukuran yang digunakan untuk menilai kualitas atau keakuratan model atau algoritma yang digunakan dalam suatu studi [18]. *Performance* dinilai berdasarkan *confusion matrix* dari berbagai metrik seperti akurasi, presisi, recall, dan lain-lain, sesuai dengan tujuan dan konsep penelitian yang telah ditetapkan [19]. Dengan ketentuan:

- TP : *true positive*
- FP : *false positive*
- FN : *false negative*
- TN : *true negative*

a. Akurasi merupakan jumlah rasio prediksi dengan jumlah total prediksi.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (5)$$

b. Presisi merupakan rasio jumlah prediksi benar dibanding jumlah total prediksi positif.

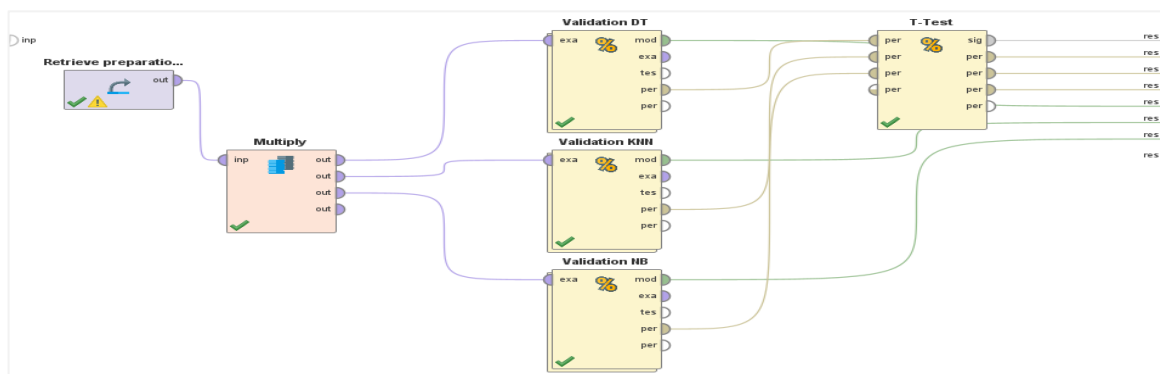
$$Precision = \frac{TP}{TP+FP} \quad (6)$$

c. Recall merupakan rasio jumlah prediksi positif dibanding dengan total prediksi benar.

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

2.7 Pengukuran Performa Data Modelling Algoritma C4.5, K-NN, dan Naïve Bayes

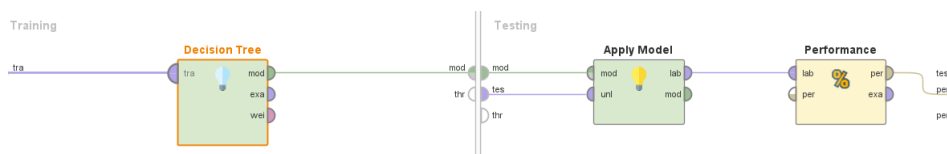
Setelah dataset diubah melalui transformasi agar sesuai, peneliti menggunakan *machine learning* untuk menggali pengetahuan dan mendapatkan akurasi prediksi menggunakan RapidMiner ver.10.3.001 sebagai *tools data mining*. Metode yang diusulkan pada penelitian kali ini merupakan perbandingan antara algoritma C4.5, K-NN, dan Naïve Bayes pada dataset pasien dengan tujuan mengukur kinerja algoritma klasifikasi mana yang lebih baik dengan hasil lebih akurat.



Gambar 2. Proses Modeling

Pada Gambar 2, merupakan proses *modeling* atau penerapan dari ketiga algoritma yaitu C4.5, K-NN, dan Naive Bayes. Diawali dengan memasukkan dataset yang telah ditransformasi kedalam *machine learning* dan memberikan label pada kolom *death_event* dengan tujuan memberi tahu model klasifikasi data mana yang harus diprediksi. Untuk melakukan validasi dataset, peneliti menggunakan metode *10-fold cross validation* dengan *sampling type automatic*. Metode ini digunakan guna membagi data menjadi 10 subset yang sama besar dengan perbandingan *0,9 data training* dan *0,1 data testing*, dengan pengujian berulang 10 kali, maka data mendapatkan performa model yang dihitung berdasarkan rata-rata hasil.

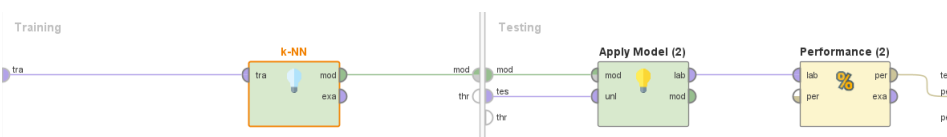
2.7 Model Klasifikasi Algoritma C4.5



Gambar 3. Parameter Algoritma C4.5

Pada gambar 3, algoritma C4.5 menggunakan kriteria *gain_ratio* yang merupakan perbandingan antara *information gain* dari suatu atribut dengan *split information* dari atribut tersebut dengan *maximal depth* sebesar 10 dan menerapkan *pruning* dan *prepruning* dengan maksud untuk menghasilkan pohon keputusan yang lebih simpel dan lebih umum, serta dapat meningkatkan kinerja pada data baru yang belum pernah terlihat sebelumnya.

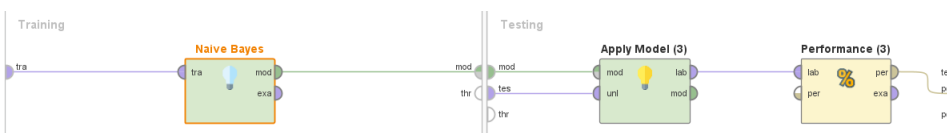
2.8 Model Klasifikasi Algoritma K-NN



Gambar 4. Parameter Algoritma K-NN

Pada gambar 4, algoritma K-NN menggunakan nilai *k* sebesar 5 dengan bantuan *weighted vote* dengan tujuan agar hasil prediksi untuk setiap titik data akan dipengaruhi oleh seberapa dekat secara jaraknya dengan tetangga terdekat, dengan bobot yang lebih besar diberikan kepada tetangga yang lebih dekat.

2.9 Model Klasifikasi Algoritma Naive Bayes



Gambar 5. Parameter Algoritma Naive Bayes

Pada gambar 5, algoritma Naive Bayes menggunakan *laplace correction* untuk menangani situasi di mana kemungkinan posterior (probabilitas kelas tertentu) menjadi nol karena tidak ada frekuensi yang diamati dalam data pelatihan. Penerapan Laplace correction melibatkan menambahkan jumlah pseudo-count ke setiap frekuensi kategori dalam model Naive Bayes. Biasanya, nilai pseudo-count yang ditambahkan adalah satu, sehingga istilah "*add-one smoothing*" sering digunakan.

3. HASIL DAN PEMBAHASAN

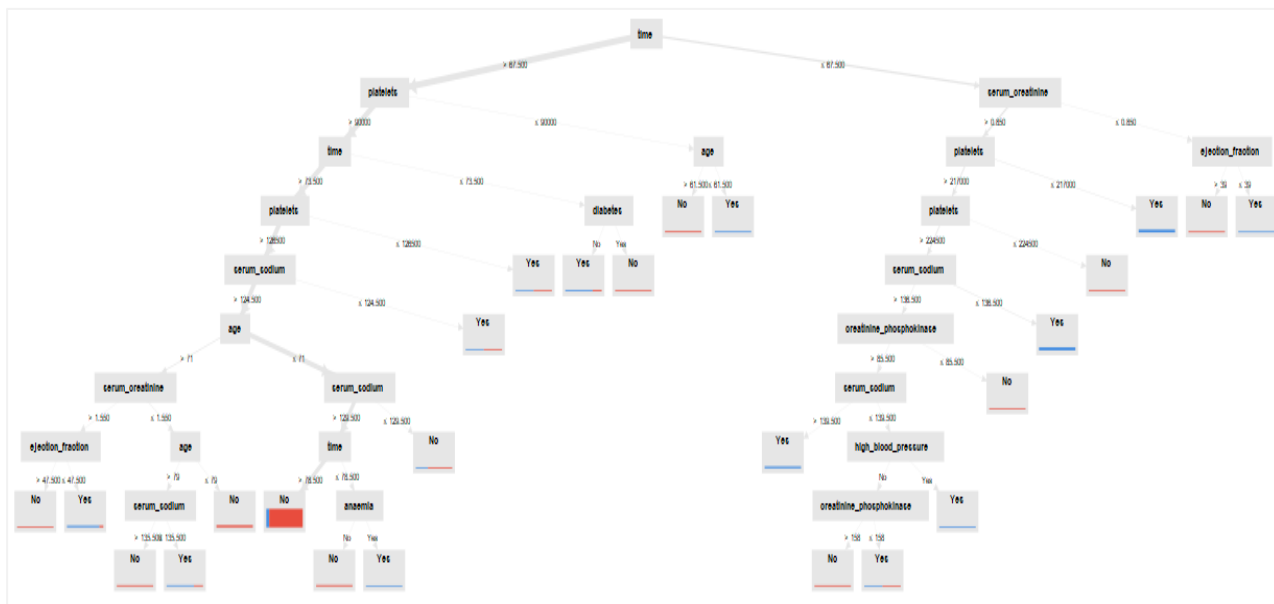
3.1 Hasil Confusion Matrix Algoritma C4.5

accuracy: 77.91% +/- 4.61% (micro average: 77.93%)

	true Yes	true No	class precision
pred. Yes	54	24	69.23%
pred. No	42	179	81.00%
class recall	56.25%	88.18%	

Gambar 6. Confusion Matrix Algoritma C4.5

Output dari algoritma C4.5 dapat dilihat pada gambar 6, dimana tabel *confusion matrix* menunjukkan akurasi prediksi sebesar 77,91% dengan standar deviasi kurang lebih 4,61% dengan total *true yes* sebanyak 54 dan *true no* sebanyak 179.



Gambar 7. Hasil dari Klasifikasi Algoritma *Decision Tree C4.5*

Pada gambar 7, algoritma C4.5 mampu menampilkan hal yang paling berpengaruh terhadap keberlangsungan hidup pasien yaitu periode penanganan medis, kemudian disusul oleh platelets atau trombosit dalam darah dan tingkat kreatinin serum dalam darah.

3.2 Hasil Confusion Matrix Algoritma K-NN

accuracy: 63.54% +/- 5.57% (micro average: 63.55%)

	true Yes	true No	class precision
pred. Yes	19	32	37.25%
pred. No	77	171	68.95%
class recall	19.79%	84.24%	

Gambar 8. Confussion Matrix Algoritma K-NN

Output dari algoritma K-NN dapat dilihat pada gambar 8, dimana tabel *confusion matrix* menunjukkan akurasi prediksi sebesar 63,54% dengan standar deviasi kurang lebih 5,57% dengan total *true yes* sebanyak 19 dan *true no* sebanyak 171.

3.3 Hasil Confusion Matrix Algoritma Naïve Bayes

accuracy: 76.59% +/- 6.85% (micro average: 76.59%)

	true Yes	true No	class precision
pred. Yes	44	18	70.97%
pred. No	52	185	78.06%
class recall	45.83%	91.13%	

Gambar 9. Confussion Matrix Algoritma Naïve Bayes

Output yang ditampilkan oleh algoritma Naïve Bayes dapat dilihat pada gambar 9, dimana tabel *confusion matrix* menunjukkan akurasi prediksi sebesar 76,59% dengan standar deviasi kurang lebih 6,85% dengan total *true yes* sebanyak 44 dan *true no* sebanyak 185.

```

SimpleDistribution
Distribution model for label attribute DEATH_EVENT

Class Yes (0.321)
12 distributions

Class No (0.679)
12 distributions
    
```

Gambar 10. Model distribusi Naïve Bayes

Pada gambar 10, diperoleh *class yes* sebesar 0,321 dan *class no* sebesar 0,679 dengan masing - masing 12 model distribusi.

3.4 Pengujian Data

Pada penelitian ini, dilakukan juga pengujian menggunakan uji *pairwise t-test* dengan tujuan menguji apakah hipotesis nol (yaitu semua nilai rata-rata aktual adalah sama) berlaku untuk vektor kinerja tertentu. Operator ini menggunakan uji-t berpasangan sederhana untuk menentukan kemungkinan hipotesis nol salah. Karena uji-t hanya dapat diterapkan pada dua vektor kinerja, pengujian ini akan diterapkan pada semua pasangan yang mungkin [20]. Hasilnya adalah matriks signifikansi. *Input* yang digunakan dalam uji *pairwise t-test* merupakan *performance vector* dengan *output t-test significance* sebagai vektor kinerja. Dalam signifikansi uji nilai statistik menggunakan tingkat menjadi 0,05 dan parameter ini menentukan ambang batas probabilitas yang menentukan apakah perbedaan dianggap signifikan.

A	B	C	D
	0.779 +/- 0.046	0.635 +/- 0.056	0.766 +/- 0.069
0.779 +/- 0.046		0.000	0.619
0.635 +/- 0.056			0.000
0.766 +/- 0.069			

Gambar 11. Hasil uji *pairwise t-test*

Gambar 11 diatas menunjukkan hasil uji bahwa algoritma dengan warna oranye, secara statistik memiliki nilai alpha kurang dari 0,05, mengidentifikasi adanya perbedaan signifikan antara nilai rata-rata. Dengan nilai akurasi C4.5 sebesar 0.779 +/- 0.046, lalu nilai akurasi K-NN sebesar 0.635 +/- 0.056, dan nilai akurasi Naive Bayes sebesar 0.766v +/- 0.069. Oleh karena itu, dapat disimpulkan bahwa terdapat perbedaan yang signifikan antara algoritma C4.5 dan K-NN, serta antara algoritma K-NN dan Naive Bayes.

4. KESIMPULAN

Berdasarkan pengolahan komparasi tiga algoritma, yaitu C4.5, K-Nearest Neighbor, dan Naive Bayes. Dengan menggunakan uji validasi *10-fold cross validation* dan *pairwise t-test*. Didapatkan hasil penelitian ini menunjukkan bahwa algoritma C4.5 merupakan algoritma dengan hasil prediksi tertinggi yaitu 0,779 dengan standar deviasi kurang lebih 0,046 dan faktor paling berpengaruh merupakan periode penanganan medis. Selanjutnya disusul oleh algoritma Naive Bayes dengan perbedaan yang tidak cukup signifikan dan terakhir algoritma K-Nearest Neighbor dengan nilai terkecil sehingga dinilai kurang baik diimplementasikan dalam dataset. Mengingat komparasi ini hanya berdasarkan tiga algoritma, penelitian selanjutnya disarankan menggunakan berbagai metode klasifikasi lainnya untuk mendapatkan hasil yang lebih optimal.

REFERENCES

- [1] E. Tasia, R. Z. I. Z. Ismail, S. K. P. Loka, Y. Ikhsani, and R. Ocviani, "Metode Klasifikasi Supervised Learning pada Penyakit Gagal Jantung: Supervised Learning Classification Method in heart failure," in SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat, 2023, pp. 1–7.
- [2] D. C. P. Buani, "Penerapan Algoritma Naive Bayes dengan Seleksi Fitur Algoritma Genetika Untuk Prediksi Gagal Jantung," *Evolusi: Jurnal Sains dan Manajemen*, vol. 9, no. 2, 2021.
- [3] D. Andri and M. Reza, "Penerapan Algoritma K-Nearest Neighbord Untuk Prediksi Kematian Akibat Penyakit Gagal Jantung," vol. III, no. 2020, pp. 105–112, 2022.
- [4] Y. Yuliani, "Algoritma Random Forest Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Seleksi Fitur Bestfirst," *Infotek J. Inform. dan Teknol.*, vol. 5, no. 2, pp. 298–306, 2022.
- [5] H. Imaduddin, B. A. Hermansyah, and M. Mutawadhi'Alfajri, "Klasifikasi Kematian Akibat Gagal Jantung Menggunakan Algoritma Logistic Regression Berbasis Forward Selection," *JIMP-Jurnal Informatika Merdeka Pasuruan*, vol. 7, no. 3, pp. 96–100, 2023.
- [6] R. Martiansah, T. A. Afifah, M. Alviyoni, and A. Arifin, "Perbandingan Algoritma Klasifikasi C4. 5 Dan Naive Bayes Untuk Memprediksi Gagal Jantung: Comparison of C4. 5 and Naive Bayes Classification Algorithms for Predicting Heart Failure," in SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat, 2023, pp. 78–85.
- [7] A. H. Nasrullah, "Implementasi algoritma Decision Tree untuk klasifikasi produk laris," *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, vol. 7, no. 2, pp. 45–51, 2021.
- [8] S. Bahri and A. Lubis, "Metode Klasifikasi Decision Tree Untuk Memprediksi Juara English Premier League," *Jurnal Sintaksis*, vol. 2, no. 1, pp. 63–70, 2020.
- [9] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi algoritma klasifikasi k-nearest neighbor (knn) untuk klasifikasi seleksi penerima beasiswa," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 6, no. 2, pp. 118–127, 2021.
- [10] A. Sefarni, I. E. Hendrawan, and C. Rozikin, "Klasifikasi Penyakit Jantung dengan Menggunakan Algoritma C4. 5," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 7, no. 2, pp. 117–126, 2022.
- [11] Q. A'yuniyah et al., "Implementasi Algoritma Naive Bayes Classifier (NBC) untuk Klasifikasi Penyakit Ginjal Kronik," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 1, pp. 72–76, 2022.

- [12] P. Subarkah, E. P. Pambudi, and S. O. N. Hidayah, "Perbandingan Metode Klasifikasi Data Mining untuk Nasabah Bank Telemarketing," *Matrik: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 1, pp. 139–148, 2020.
- [13] S. D. Damanik and M. I. Jambak, "Klasifikasi Customer Churn pada Telekomunikasi Industri Untuk Retensi Pelanggan Menggunakan Algoritma C4.5," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 6, pp. 1303–1309, 2023.
- [14] F. A. Pambudi, A. P. Windarto, M. Fauzan, J. T. Hardinata, and R. Winanjaya, "Analisis Klasifikasi C4.5 Pada Pola Pembayaran Sepeda Motor Adira Cabang Pematangsiantar," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 1, no. 4, pp. 160–169, 2021.
- [15] K. Kusnawi and K. I. F. Pratama, "Komparasi Algoritma Supervised Learning dan Feature Selection pada Klasifikasi Penyakit Gagal Jantung," *Indonesian Journal of Computer Science*, vol. 12, no. 6, 2023.
- [16] A. Tangkelayuk, "The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree," *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, vol. 9, no. 2, pp. 1109–1119, 2022.
- [17] M. R. Yuliansyah, B. Muslimin, and A. Franz, "Perbandingan Metode K-Nearest Neighbors dan Naïve Bayes Classifier Pada Klasifikasi Status Gizi Balita di Puskesmas Muara Jawa Kota Samarinda," *Adopsi Teknologi dan Sistem Informasi (ATASI)*, vol. 1, no. 1, pp. 8–20, 2022.
- [18] A. Putri et al., "Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir: Comparison of K-NN, Naive Bayes and SVM Algorithms for Final-Year Student Graduation Prediction," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, pp. 20–26, 2023.
- [19] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Riset Komputer)*, vol. 8, no. 6, pp. 219–225, 2021.
- [20] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," *Indonesian Journal of Data and Science*, vol. 1, no. 2, pp. 29–33, 2020.