

PERBANDINGAN METODE Naïve BAYES DENGAN DECISION TREE PADA KLASIFIKASI DATA PASIEN YANG TERKENA SERANGAN JANTUNG

*(Method Comparison Of Naïve Bayes And Decision Tree To The Data Classification Of
Patients With Heart Attack)*

Anggun Dwi Rizkika^[1], Mustika Dewi Waskita^[2], Nurun Latifah^[3], Pande Permadi Kusuma^[4]

^[1]Dept Informatics Engineering, Mataram University
Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA

Email: anggundwrzka@gmail.com, mustikadewi990@gmail.com, latifahnurun08@gmail.com,
ppermadik195@gmail.com

Abstract

Heart attack is one of the deadliest diseases in the world and is the highest cause of death after stroke in Indonesia. Technological advances in the field of data processing are one way to help research on heart disease. Classification methods with certain algorithms, such as Naïve Bayes and Decision Tree are popular ways to predict data. The Naïve Bayes method uses probabilities in each data, while the Decision Tree method uses a decision tree. This paper aims to compare the two methods in classifying patient data with heart disease. The comparison of methods will be seen from the performance measure consisting of recall, precision, and accuracy of each class. Based on the classification results of 303 patient data with each method, it is found that the Naïve Bayes algorithm is more optimal than the Decision Tree algorithm. Naïve Bayes algorithm has an accuracy of 88.54% with a recall of 90.32% and a precision of 87.5%. While the Decision Tree algorithm has an accuracy of 78.69% with a recall result of 82.76% and a precision of 75%.

Keywords: Classification, Machine Learning, Heart Attack Disease, Naïve Bayes, Decision Tree

1. PENDAHULUAN

Kesehatan sangat penting untuk dijaga oleh setiap manusia agar dapat melakukan aktivitas kehidupan sehari-hari. Manfaat menjaga kesehatan agar tidak mudah lelah dan lebih berfokus dalam menyelesaikan rutinitas keseharian bagi manusia. Menjaga kesehatan juga berdampak di dalam tubuh manusia karena terdapat beberapa organ vital, salah satunya jantung. Jantung menjadi organ yang berperan penting sebagai alat pemompa darah yang untuk mengalirkan darah ke seluruh tubuh manusia. Apabila terjadi gangguan atau kerusakan pada organ jantung maka dapat mengakibatkan kinerja dari seluruh organ di dalam tubuh manusia dapat terganggu[1].

Penyakit jantung merupakan penyakit yang serius dan dapat mematikan. Menurut WHO, lebih dari 7 juta orang di seluruh dunia meninggal akibat penyakit jantung pada tahun 2002, dan angka tersebut diperkirakan akan meningkat hingga 11 juta orang pada tahun 2020. Di Indonesia, hasil survey Sample Registration System (SRS) pada tahun 2014

menunjukkan bahwa penyakit jantung merupakan penyebab kematian tertinggi setelah stroke, dengan persentase sebesar 12,9%. Berdasarkan data dari GLOBOCAN pada tahun 2012, terdapat 43,30% kasus serangan jantung dan 12,90% kematian akibat penyakit jantung di Indonesia. Pada tahun 2013, terdapat sebanyak 61.682 orang yang menderita penyakit jantung, dengan provinsi Jawa Tengah memiliki jumlah kasus terbanyak yaitu 11.511 kasus [2].

Pola hidup yang kurang teratur, makanan yang tidak dijaga dan kurangnya informasi terhadap penyakit jantung ini dapat menyebabkan seseorang menjadi lengah sehingga tidak dapat mengenali gejala awal dari serangan jantung. Pendeteksian penyakit jantung secara manual dapat dilakukan, yaitu dengan melakukan konsultasi atau berobat ke dokter dan melakukan beberapa pemeriksaan di laboratorium yang selanjutnya harus dikonsultasikan kembali dengan dokter spesialis jantung. Hal ini tentu saja membuat pasien mengeluarkan biaya yang cukup mahal. Terdapat hubungan antara akses informasi

yang kurang atau keterlambatan media untuk pemeriksaan awal. Akses yang minim untuk masyarakat dapat mencari informasi tentang penyakit serangan jantung ini dapat mengakibatkan terjadinya peningkatan angka kematian setiap tahunnya.

Tingginya risiko kematian akibat dari terjadinya serangan jantung membuat perlunya dibangun sebuah sistem yang mampu melakukan pendeteksian dini potensi serangan jantung pada pasien secara akurat dengan biaya yang terjangkau. Deteksi lebih awal butuh dilakukan dengan cara mengumpulkan data rekam medis pasien. Untuk merealisasikannya, diperlukan metode yang dapat mengolah data tersebut. Salah satu cara untuk mendeteksi penyakit tersebut adalah penggunaan suatu teknik dari data mining, yaitu klasifikasi.

Dalam klasifikasi, dilakukan pengolahan terlebih dahulu dengan data yang diperoleh sebelumnya berdasarkan variabel yang ada untuk memisahkan perbedaan kategori. Untuk penelitian ini, penulis memanfaatkan Algoritma Naïve Bayes dan Decision Tree yang selanjutnya dibandingkan melalui variabel yang berpengaruh pada hasil akurasi.

Berdasarkan pemaparan sebelumnya, maka batasan masalah yang akan ditekankan dalam penelitian ini adalah objek penelitian berupa data pasien terkena serangan jantung. Penelitian ini diharapkan dapat mengungkapkan di antara metode Naïve Bayes dan Decision Tree yang paling akurat dalam melakukan klasifikasi guna deteksi awal penyakit jantung.

2. TINJAUAN PUSTAKA

Dalam proses perancangan dan implementasi penelitian, terdapat beberapa penelitian terkait yang relevan dan dapat dijadikan referensi pendukung pada penelitian sebagai berikut:

Terdapat beberapa penelitian untuk memprediksi penyakit jantung, diantaranya dilakukan oleh Amril Samosir, Ms Hasibuan, Wahyu Eko dan Tri Hariyono. dengan melakukan komparasi terhadap 3 algoritma yaitu *Random Forest*, *Naïve Bayes*, dan *K-Nearest Neighbor*. Hasil dari penelitian tersebut klasifikasi algoritma *Naïve Bayes* menghasilkan performance measure yang terbaik untuk mengklasifikasi penyakit jantung dengan nilai AUC 0,910 dibandingkan metode klasifikasi algoritma *Random Forest* dan *K-Nearest Neighbor* dengan masing masing nilai AUC 0,899 dan 0,686 [2].

Pada penelitian lain dengan judul Algoritma Decision Tree dan Smote Untuk Klasifikasi Serangan

Jantung Miokarditis yang Imbalance. Algoritma yang akan dievaluasi yaitu *Naïve Bayes*, *Support Vector Machine*, dan *Decision Tree*. Penelitian ini menggunakan dataset *Myocardial Infarction Complications* yaitu data komplikasi pasien yang telah di *preprocessing*. Sehingga diperoleh hasil algoritma Decision Tree dengan akurasi 98%, diikuti algoritma Support Vector Machine dengan akurasi 91% dan Naïve Bayes dengan akurasi paling rendah yakni 49% [3].

Pada topik penelitian yang berbeda yaitu Perbandingan Klasifikasi Metode *Naïve Bayes* dan *Decision Tree* Algoritma (J48) Pada Pasien Penderita *Stroke* di RSUD Abdul Wahab Sjahranie Samarinda. Penelitian ini menggunakan data *training* sebesar 90% data dari 156 dataset. Hasil akurasi metode *Decision Tree* lebih besar yaitu 87,5% dibandingkan metode *Naïve Bayes* dengan akurasi 81,25% [4].

Berdasarkan tinjauan pustaka dari penelitian-penelitian terdahulu yang telah dilakukan, peneliti memilih metode *Naïve Bayes* dan *Decision Tree* untuk mengetahui akurasi data. Pada penelitian ini, dilakukan suatu percobaan untuk membandingkan tingkat akurasi dari metode Algoritma *Naïve Bayes* dan *Decision Tree* dalam klasifikasi penyakit Serangan Jantung.

2.1. Klasifikasi

Klasifikasi merupakan sebuah proses untuk melakukan penilaian objek data untuk memasukkannya dalam kelas tertentu dari beberapa kelas yang tersedia. Dalam proses klasifikasi data terdapat dua tahap utama yang akan dilakukan, yaitu tahap pertama adalah membangun model yang menjadi prototipe untuk disimpan menjadi memori atau pengetahuan dan tahap kedua adalah model yang telah dibuat akan digunakan untuk proses klasifikasi atau mengenali ataupun memprediksi suatu objek atau data baru untuk diketahui objek atau data baru tersebut masuk ke dalam kelas yang mana dari model yang sudah dibangun sebelumnya [5].

Dalam proses klasifikasi, dataset akan dibagi kedalam 2 bagian, yaitu data latih (data *training*) dan data uji (data *testing*), dimana keseluruhan dataset akan dibagi menjadi dua cluster dan pada data tersebut akan dilakukan proses pelatihan dan pengujian. Data yang telah dibagi tersebut akan menentukan hasil dari akurasi penggunaan metode dalam proses klasifikasi tersebut [5].

2.2. Naïve Bayes

Naïve Bayes merupakan salah satu metode untuk melakukan proses klasifikasi probabilistik sederhana. Teorema dari Naive bayes ini diusulkan oleh seorang

ilmuwan Inggris yaitu Thomas Bayes yang melakukan prediksi probabilitas di masa depan berdasarkan pada pengalaman yang ada di masa lalu. Metode ini bekerja dengan melakukan penghitungan dari sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Teorema Bayes dijadikan sebagai dasar dalam penggunaan Naïve Bayes, dimana teorema ini mempunyai kemampuan klasifikasi serupa dengan decision tree dan neural network. Teorema Bayes memiliki bentuk umum seperti:

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

Naïve Bayes merupakan metode yang terbentuk dari penyederhanaan Teorema Bayes. Berikut ini merupakan rumus Naïve Bayes setelah disederhanakan:

$$P(X|H) = P(H|X) P(X)$$

Keterangan:

H : hipotesis data X, merupakan suatu class spesifik

$P(H|X)$: probabilitas hipotesis H berdasarkan kondisi X (posteriori probability)

$P(H)$: probabilitas hipotesis H (prior probability)

$P(X|H)$: probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: probabilitas dari X

Kelebihan dari penggunaan Naive Bayes sebagai metode klasifikasi adalah data yang dibutuhkan lebih sedikit untuk digunakan sebagai data pelatihan untuk parameter pengklasifikasi. Karena variabel independen diasumsikan, hanya variasi variabel untuk setiap kelas yang perlu ditentukan, bukan keseluruhan matriks kovarians. Dengan demikian, Naive Bayes mengasumsikan bahwa ada atau tidak adanya atribut di kelas tidak bergantung pada ada atau tidak adanya atribut lain di kelas yang sama. Saat mengklasifikasikan, metode Bayesian akan menghasilkan label kelas dengan probabilitas tertinggi [1]. Selain itu, kelebihan dari menggunakan metode Naïve Bayes adalah dengan algoritma yang sederhana namun masalah dapat diselesaikan dengan cepat dengan akurasi yang tinggi [6].

2.3. Decision Tree

Decision tree merupakan suatu diagram alir dengan struktur yang mirip seperti pohon, dimana untuk atribut-atribut yang dilakukan pengujian akan dinotasikan pada interval node, hasil dari atribut-atribut tersebut akan direpresentasikan oleh setiap cabang dari tree, dan kelas-kelas atau distribusi

dari setiap kelas direpresentasikan pada leaf node atau simpul daun [7].

Tahap klasifikasi dengan decision tree akan terhenti pada suatu node atau simpul apabila kelas atau jenis objek sudah ditemukan pada node tersebut. Apabila suatu objek sudah ditemukan dalam satu tingkat maka objek tersebut telah diketahui termasuk ke dalam kelas tertentu, namun jika tidak maka akan berlanjut dengan diberikannya pertanyaan pada level berikutnya hingga ciri-cirinya diperoleh dengan jelas dan jenis dari objek dapat ditentukan [7].

Penerapan dari algoritma decision tree kini telah melengkapi kebutuhan untuk dapat melakukan pemetaan terhadap faktor-faktor pendukung sehingga dapat dikategorikan sebagai suspek penyakit serangan jantung pada manusia. Faktor-faktor penyebab serangan jantung yang berhasil diketahui diharapkan dapat membawa ke hasil akhir untuk dapat dilakukan tindak lanjut pencegahan agar peluang faktor risiko kejadian serangan jantung yang tinggi dapat dihindari.

Secara umum algoritma Decision Tree adalah untuk membangun pohon keputusan sebagai berikut :

- Pilih suatu atribut untuk dijadikan akar
- Buat cabang untuk setiap nilai yang ada
- Bagi kasus dalam cabang
- Ulangi proses untuk masing-masing cabang hingga seluruh kasus pada cabang memiliki kelas yang sama.

Rumus menghitung nilai entropy menggunakan persamaan [8]

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

Keterangan :

S = himpunan kasus

n = jumlah partisi atribut A

P_i = proporsi S_i terhadap S

$|S_i|$ = jumlah kasus pada partisi ke i

$|S|$ = jumlah kasus dalam S

A = atribut

Rumus untuk mencari nilai gain :

$$Gain(S,A) = \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy$$

2.4. Pengukuran Akurasi

a. Confusion Matrix

Pengukuran dari sebuah metode salah satunya dilakukan dengan menggunakan tabulasi silang. Berisi informasi terkait kelas data asli yang direpresentasikan pada baris matriks dan kelas data hasil prediksi suatu algoritma direpresentasikan pada kolom klasifikasi [6].

TABEL I. CONFUSION MATRIX

	<i>Classified Positive</i>	<i>Classified Negative</i>
<i>Actual Positive</i>	TP	FN
<i>Actual Negative</i>	FP	TN

b. *Recall*

Recall merupakan istilah yang digunakan untuk dokumen terpanggil yang relevan dengan pernyataan (*query*) yang dimasukkan pengguna dalam suatu sistem temu balik informasi. *Recall* berhubungan dengan kemampuan suatu sistem temu balik dalam menemukan dokumen yang relevan. Hal ini berarti *recall* adalah bagian dari proses temu balik informasi yang dapat digunakan sebagai alat ukur tingkat efektivitas suatu sistem temu balik informasi [6].

$$Recall = \frac{TP}{TP + FN}$$

c. *Precision*

Precision merupakan jumlah dari kelompok dokumen relevan dari total jumlah dokumen yang didapatkan oleh sistem. Presisi juga menjadi sebuah cara untuk mengukur tingkat efektivitas sistem temu balik informasi. Pengukuran tingkat ketepatan dalam kegiatan penelusuran Jumlah dokumen relevan yang terambil [6].

$$Precision = \frac{TP}{TP + FP}$$

d. *Accuracy*

Accuracy didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual [6].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

3. METODE PENELITIAN

3.1 Alat dan Bahan

Alat dan bahan yang dibutuhkan pada penelitian ini berupa software, hardware serta data dan informasi pendukung selama dilakukannya penelitian.

a. Alat

Berikut merupakan alat-alat berupa software dan hardware yang digunakan di dalam pembuatan artikel penelitian:

- Laptop dengan spesifikasi processor AMD Ryzen 7 RAM 8 GB sebagai sarana untuk melakukan penelitian dan pengembangan sistem.
 - Sistem Operasi Windows 10
 - Microsoft Office Word 2010 untuk penyusunan artikel penelitian
 - *Software* Chrome untuk mencari referensi
 - Visual Studio Code, sebagai *text editor* pada saat melakukan proses *coding*.
- b. Bahan

Bahan-bahan yang digunakan untuk penelitian ini yaitu, dataset dari situs kaggle.com yang berupa file .csv bernama heart.csv. yang terdiri dari 303 baris dan 14 kolom.

Dataset ini dibagi menjadi atribut independen atau atribut yang mempengaruhi dan atribut dependen atau atribut yang dipengaruhi. Atribut independen berupa *Age*, *sex*, *chest pain (cp)*, *resting blood pressure (trtbps)*, *cholesterol (chol)*, *fasting blood sugar (fbs)*, *resting electrocardiographic (restecg)*, *thalachh*, *exercise induced angina (exng)*, *oldpeak*, *slope (slp)*, *ca*, dan *thall*. Sedangkan atribut dependen yaitu *output*.

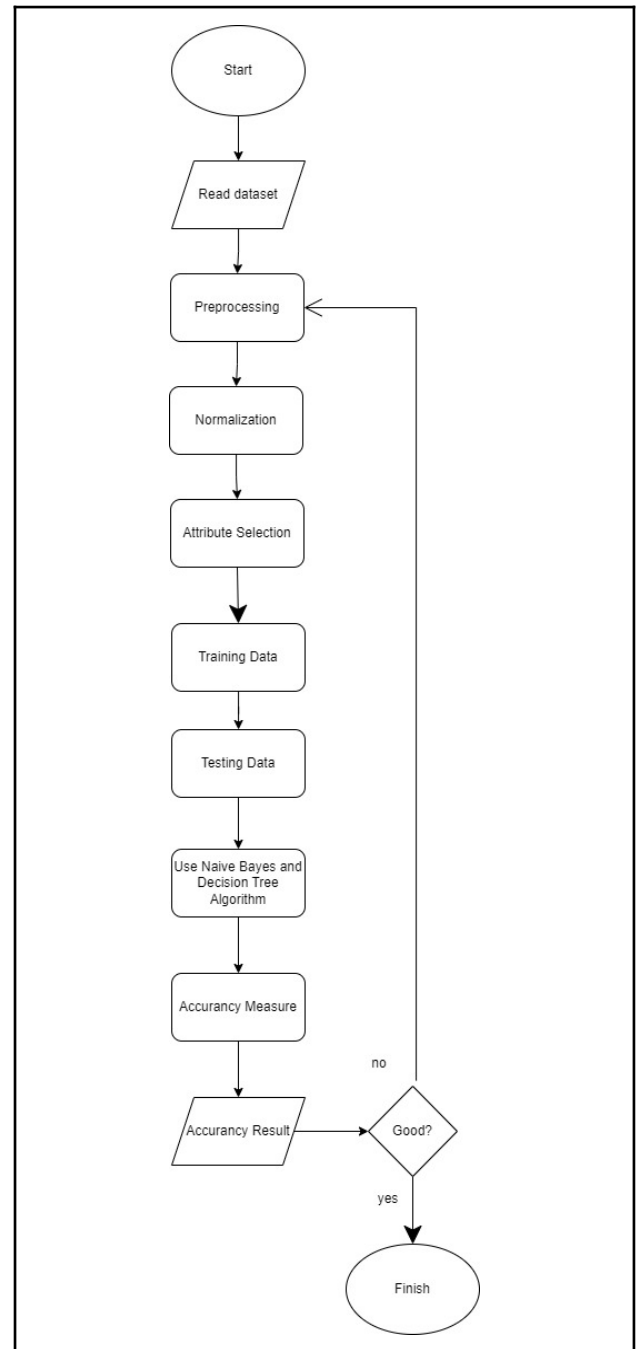
3.2 Pengumpulan Data

TABEL II. VARIABEL DATA

No	Variabel	Keterangan
1	Age	Usia pasien
2	Sex	Jenis kelamin pasien (1=Laki-Laki, 0=Perempuan)
3	Cp	Tipe nyeri dada (0=asymptomatic, 1=atypical angina, 2=non anginal pain, 3=typical anginal)
4	Trtbps	Tekanan darah pasien
5	Chol	Kolesterol pasien
6	Fbs	Kadar gula darah
7	Restecg	Kondisi ECG pasien dalam keadaan istirahat
8	Thalachh	Rata-rata detak jantung pasien dalam 1 menit.
9	Exng	Nyeri dada pasien (1=yes, 0=no)

10	Oldpeak	Penurunan ST setelah berolahraga
11	Slp	Slope puncak ST setelah berolahraga (0= <i>downsloping</i> , 1= <i>flat</i> , 2= <i>upsloping</i>)
12	Caa	Jumlah pembuluh darah pada proses <i>flourosopy</i>
13	Thall	Detak jantung Pasien
14	<i>Output</i>	(0=tidak beresiko, 1=beresiko)

3.3. Tahapan Penelitian



Gambar 1. Flowchart

Penelitian ini dilakukan beberapa proses dengan alur yang tentunya dimulai dengan mengumpulkan data yang akan digunakan kemudian dilakukan proses coding yang mana data yang telah dikumpulkan dibaca dan dilakukan proses *preprocessing* dataset, kemudian dilakukan proses normalisasi, setelah keseluruhan data normal, dilakukan proses test dan training untuk memperoleh hasil presisi, recall, dan akurasi dari data.

3.4. Pengolahan Data

- Preprocessing* dengan melakukan pengecekan dataset untuk melihat apakah terdapat data yang tidak seragam, tidak sesuai, ataupun terdapat nilai yang tidak lengkap.
- Normalisasi pada data dilakukan satu persatu pada setiap kolom. Data dapat dikatakan normal jika tidak memiliki perbedaan yang signifikan, serta data berada pada kelas yang sama. Sehingga dapat memberikan hasil terbaik.
- Menentukan atribut-atribut yang akan digunakan untuk klasifikasi berdasarkan nilai korelasinya dengan atribut *output*. Atribut yang digunakan yaitu, *cp*, *thalach*, *exang*, *oldspcak*, *slp*, *caa*, *thall*, dan *output*.
- Membagi data *training* dan data *testing* yaitu mengambil 20% dari keseluruhan data untuk dilakukan *testing* dan 80% di-*training*. Data *training* merupakan bagian dataset yang berisi kumpulan data yang akan dilatih agar prediksi dapat dibuat dan dapat membangun model prediksi. Sedangkan data *testing* merupakan bagian dataset yang berisi kumpulan data yang akan diuji atau dites untuk mengetahui seberapa besar keakuratan ataupun performanya. Dataset dibagi menjadi data *input* yang mana data akan diberi label "x" dan *output* diberi label "y".
- Melakukan *test* dan *training* data dengan menggunakan metode *Naïve bayes* dan *Decision tree*.
- Menampilkan hasil berupa tingkat akurasi data dengan menggunakan kedua data tersebut.

3.5 Model yang Diusulkan

Dalam penelitian ini, metode yang diusulkan dalam proses pengolahan dan pengujian data menggunakan Algoritma *Naïve bayes* dan *Decision tree* untuk memprediksi dan menentukan tingkat akurasi data.

3.6 Evaluasi dan Validasi Hasil

Pada penelitian ini, dilakukan evaluasi terhadap model yang ditentukan untuk mengetahui tingkat keakuratan model.

4. HASIL DAN PEMBAHASAN

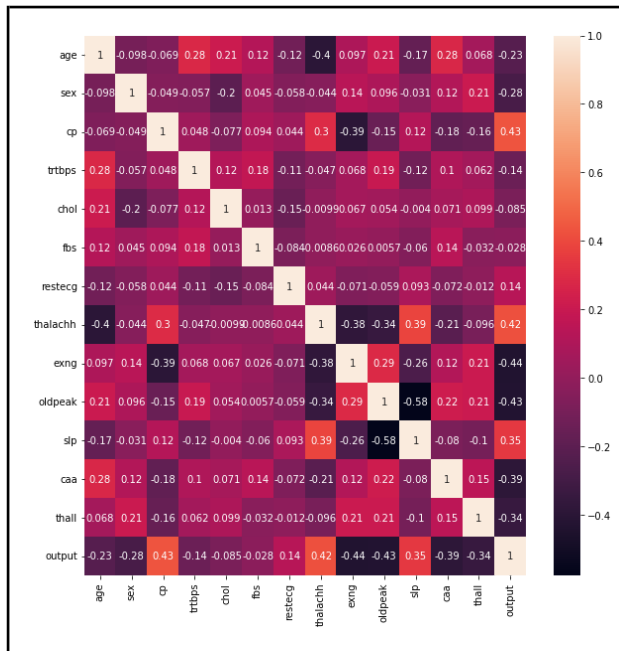
Penelitian dilakukan dengan tujuan untuk membandingkan kinerja metode klasifikasi dan mendapatkan hasil akurasi yang baik pada klasifikasi data pasien yang terkena penyakit jantung menggunakan metode *Naïve Bayes* dan *Decision Tree*. Penelitian ini menggunakan 303 data pasien dengan total variabel sejumlah 14.

Data yang sudah di-*import* kemudian memasuki tahap awal pemrosesan, yaitu dilakukannya proses *preprocessing* data. Data mentah yang diambil akan di *preprocessing* agar menjadi format yang lebih mudah dan efektif untuk diproses sehingga pengolahan data dan pengujian dapat mengeluarkan hasil yang lebih akurat. *Preprocessing* data dilakukan dengan mengecek apakah dalam dataset terdapat nilai yang kosong. Apabila ditemukan data yang tidak lengkap atau nilai yang kosong di dalam tabel dataset maka baris yang menjadi letak data yang kosong tersebut akan dihapus atau dihilangkan. Pada dataset *heart.csv* setelah dilakukan *preprocessing* didapatkan hasil bahwa tidak ditemukannya data yang kosong. Artinya dataset tersebut sudah seragam dan tidak terdapat data yang kosong. Sehingga tidak terdapat perubahan jumlah baris dan kolom data sebelum dan setelah *preprocessing*.

Selanjutnya dilakukan pengecekan terhadap keseimbangan distribusi data di dalam dataset. Pengecekan dilakukan dengan melihat distribusi data pada kolom *output* dalam dataset. Kolom *output* merupakan kesimpulan yang terbagi menjadi "0" untuk tidak berisiko terkena serangan jantung dan "1" untuk berisiko terkena serangan jantung. Hasil yang diperoleh dari pengecekan distribusi data pada 303 baris data adalah didapatkan jumlah dari kelompok "0" adalah 138 data dan pada kelompok "1" berjumlah 165 data. Jumlah data tersebut tidak jauh berbeda sehingga dapat dikatakan bahwa dataset *heart.csv* merupakan dataset yang seimbang (*balanced*) dalam hal distribusi datanya.

Data akan dinormalisasi sehingga tiap data pada kolom mempunyai nilai dengan rentang antara 0 dan 1. Normalisasi pada data dilakukan agar data dapat diolah dengan lebih cepat.

Untuk klasifikasi menggunakan *Naïve Bayes* dan *Decision Tree*, dibuatkan plot dari nilai korelasi datanya sehingga pemilihan atribut yang digunakan dapat lebih mudah dilakukan. Nilai korelasi merupakan suatu ukuran data yang bertujuan untuk melihat hubungan antara dimensi data, sehingga dapat dilihat hubungan antara tiap atribut.



Gambar 2. Plot Heatmap

Setelah melihat nilai korelasi tiap atribut pada heatmap, ditentukan bahwa atribut yang digunakan adalah atribut yang mempunyai nilai korelasi kurang dari -0.3 dan lebih dari 0.3 dengan atribut 'output'. Sehingga atribut yang digunakan adalah cp, thalach, exang, oldpeak, slp, caa, thall, dan output.

Pada proses klasifikasi, digunakan 3 jenis dataset yang sudah dibersihkan dan dinormalisasi, yaitu dataset yang terdiri dari seluruh fitur, dataset dengan 5 fitur yang memiliki nilai korelasi lebih dari 0.4 dan kurang dari -0.4, serta dataset dengan 8 fitur yang memiliki nilai korelasi lebih dari 0.3 dan kurang dari -0.3. Hasil klasifikasi dengan menggunakan algoritma Naïve Bayes dapat dilihat pada tabel di bawah ini.

TABEL III. HASIL KLASIFIKASI DENGAN METODE NAÏVE BAYES

Fitur	Akurasi (%)	Recall (%)	Presisi (%)
'age', 'sex', 'cp', 'trtbps', 'chol', 'fbs', 'restecg', 'thalachh', 'exng', 'oldpeak', 'slp', 'caa', 'thall', 'output' (14 fitur)	88,52	90,32	87,5
'cp', 'thalachh', 'exng', 'oldpeak', 'output' (5 fitur)	83,60	84,38	84,35
'cp', 'thalachh', 'exng', 'oldpeak', 'slp', 'caa', 'thall', 'output' (8 fitur)	90,16	93,33	87,5

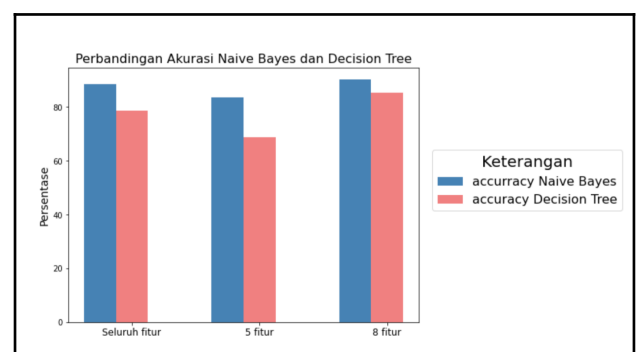
Berdasarkan Tabel III, didapat bahwa klasifikasi menggunakan algoritma Naïve Bayes menghasilkan nilai akurasi, recall, dan presisi terbesar pada dataset dengan 8 fitur, yaitu sebesar 90,16%, 93,33%, dan 87,5%. Pada tabel tersebut terlihat bahwa nilai akurasi, recall, dan presisi paling kecil terdapat pada dataset dengan 5 fitur, yaitu sebesar 83,60%, 84,38%, dan 84,35%.

TABEL IV. HASIL KLASIFIKASI DENGAN METODE DECISION TREE

Fitur	Akurasi (%)	Recall (%)	Presisi (%)
'age', 'sex', 'cp', 'trtbps', 'chol', 'fbs', 'restecg', 'thalachh', 'exng', 'oldpeak', 'slp', 'caa', 'thall', 'output' (14 fitur)	78,69	82,76	75
'cp', 'thalachh', 'exng', 'oldpeak', 'output' (5 fitur)	68,86	69,69	71,88
'cp', 'thalachh', 'exng', 'oldpeak', 'slp', 'caa', 'thall', 'output' (8 fitur)	85,25	92,59	78,13

Berdasarkan Tabel IV, didapat bahwa klasifikasi menggunakan algoritma Decision Tree menghasilkan nilai akurasi, recall, dan presisi terbesar pada dataset dengan 8 fitur, yaitu sebesar 82,25%, 92,59%, dan 78,13%. Pada tabel tersebut terlihat bahwa nilai akurasi, recall, dan presisi paling kecil terdapat pada dataset dengan 5 fitur, yaitu sebesar 68,86%, 69,69%, dan 71,88%.

Perbandingan hasil akurasi yang diperoleh dengan algoritma Naive Bayes dan Decision tree ditunjukkan dengan visualisasi grafik berikut.



Gambar 3. Perbandingan Akurasi Kedua Metode

Berdasarkan grafik pada Gambar 3. dapat dilihat bahwa terjadi peningkatan akurasi akibat dari seleksi fitur. Penggunaan dataset dengan 8 fitur menghasilkan tingkat akurasi yang lebih tinggi seperti yang ditunjukkan pada grafik tersebut. Selain itu perbedaan dari perbandingan antara akurasi Naive Bayes dengan Decision Tree juga ditunjukkan melalui visualisasi grafik tersebut yang mana tingkat akurasi dari algoritma Naive Bayes memperoleh hasil yang lebih tinggi dibandingkan dengan akurasi dari algoritma Decision Tree.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan pada analisis, dapat disimpulkan bahwa hasil akurasi terbesar klasifikasi penyakit serangan jantung dengan metode *Naive bayes* adalah 90,16% dan metode *decision tree* diperoleh tingkat akurasi terbesar sebesar 85,25%. Hal ini menunjukkan bahwa pada penelitian ini, metode *Naive bayes* memberikan ketepatan prediksi klasifikasi yang lebih baik dibanding metode *decision tree*. Didapatkan juga kesimpulan bahwa pemilihan atribut yang tepat dengan jumlah yang tepat pula, dapat menghasilkan nilai akurasi yang lebih tinggi pada tiap metode.

5.2 Saran

Saran untuk penelitian selanjutnya diharapkan untuk untuk menambah metode guna memperoleh hasil yang lebih baik serta diharapkan menggunakan lebih banyak sumber untuk mencari topik yang dapat dijadikan bahan penelitian.

UCAPAN TERIMA KASIH

Segala puji dan syukur peneliti panjatkan kehadirat Tuhan Yang Maha Esa. Karena berkat, rahmat dan karunia serta mukjizatNya, sehingga peneliti dapat menyelesaikan penyusunan artikel ilmiah ini.

Pada proses penelitian dan penyusunan artikel ini, peneliti menyadari betul bahwa ada orang-orang yang membantu penelitian dan penyusunan, di antaranya untuk:

1. Dr. Ario Yudo Husodo S.T. M.T. selaku dosen pengampu mata kuliah Kecerdasan Buatan yang telah membimbing dan memberikan penjelasan materi selama masa perkuliahan
2. Orangtua yang senantiasa memberikan doa, restu, dan dukungan serta motivasi untuk kami sehingga kami tetap semangat dalam menuntut ilmu sampai saat ini.

3. Teman-teman dalam kelompok tugas besar yang telah bekerja sama dengan baik hingga akhir untuk menuntaskan artikel ilmiah ini.

DAFTAR PUSTAKA

- [1] A. Samosir, M. Hasibuan, W.E. Justino, T.Hariyono, "Komparasi Algoritma Random Forest, Naïve Bayes dan KNearest Neighbor Dalam klasifikasi Data Penyakit Jantung," Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2021, vol 1, hal 214-222, 2021.
- [2] H. Setiani, A. Sunyoto, dan A. Nasiri, "Metode Naïve Bayes dan Particle Swarm Optimization untuk Klasifikasi Penyakit Jantung," *EXPLORE Universitas Amikom Yogyakarta*, vol. 12, no.2, hal.6-12, 2022.
- [3] A. F. Putri, A. Saifuddin, N. A. Syafaz, N. Yudistira, "Algoritma Decision Tree Dan Smote Untuk Klasifikasi Serangan Jantung Miokarditis Yang Imbalance," *JurnalLitbangEdusaintech (JLE)*, vol. 2. no. 2. hal. 112-122, 2021.
- [4] I. Lishania, R. Goejantoro, dan Y. N. Nasution, "Perbandingan Klasifikasi Metode Naïve Bayes dan Metode Decision Tree Algoritma (J48) pada Pasien Penderita Penyakit Stroke di RSUD Abdul Wahab Sjahranie Samarinda," *Jurnal EKSPONENSIAL*, vol. 10, no. 2, hal. 135-142, 2019.
- [5] E. Prasetyo, *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*, Yogyakarta: Andi Offset, 2012.
- [6] M. A. Bianto, K. Kusriani, and S. Sudarmawan, "Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes," *Creative Information Technology Journal*, vol. 6, no. 1, pp. 75–83, 2020.
- [7] S. Islamiati, I. M. Widiartha, "Klasifikasi Penyakit Jantung Menggunakan Metode Decision Tree Dengan Penerapan Algoritma C5.0," *Prosiding Seminar Nasional Teknologi Infotmasi & Aplikasinya*, hal. 308-316, 2015.
- [8] B. A. C. Permana, I. K. Dewi, "Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes." *Infotek : Jurnal Informatika dan Teknologi*, vol. 4, no.1, hal 63-69, 2021.