

## **PENERAPAN ALGORITMA NAIVE BAYES UNTUK PREDIKSI PENYAKIT JANTUNG**

Ananda Mutia Dewi<sup>1)</sup>, Arista Ardilla<sup>2)</sup>, Esar Alkausar<sup>3)</sup>, Dian Yasmine<sup>1,2,3,4)</sup>  
Prodi Informatika Medis Fakultas Kesehatan Teknologi dan Sains Universitas Bumi  
Persada

\*Correspondence: [anandamutia19@gmail.com](mailto:anandamutia19@gmail.com)

### **ABSTRAK:**

Latar Belakang: Perkembangan ilmu pengetahuan telah merambah berbagai bidang, termasuk kesehatan, di mana ilmu pengetahuan digunakan untuk pengolahan data. Salah satu contohnya adalah pengolahan data pasien penyakit jantung. Namun, sebagian besar data pasien tersebut belum dimanfaatkan secara optimal, sehingga informasi yang dihasilkan kurang efektif untuk mendukung pengambilan keputusan terkait kesehatan pasien, baik yang menderita penyakit jantung maupun tidak. Pengambilan keputusan yang didasarkan pada data atau informasi yang akurat dapat menghasilkan prediksi yang lebih tepat mengenai kondisi kesehatan pasien. Tujuan Penelitian: Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan sebuah sistem prediksi penyakit jantung menggunakan data historis. Sistem ini dirancang untuk membantu proses pengambilan keputusan secara lebih akurat dengan memanfaatkan algoritma Naive Bayes dan bahasa pemrograman Python. Metode Penelitian: Penelitian ini menggunakan metode kuantitatif dengan pendekatan deskriptif. Algoritma Naive Bayes diterapkan untuk menganalisis dataset penyakit jantung guna menghasilkan prediksi yang akurat. Hasil Penelitian: Hasil penelitian menunjukkan bahwa sistem prediksi yang dibangun mampu mengidentifikasi penyakit jantung dengan tingkat akurasi yang cukup tinggi, yaitu 82%, 83%, dan 84%. Pengujian dilakukan menggunakan dataset yang terdiri dari 1.024 data pasien. Kesimpulan: Dapat disimpulkan bahwa sistem prediksi berbasis algoritma Naive Bayes ini efektif dalam memprediksi penyakit jantung dengan akurasi yang baik. Sistem ini dapat menjadi alat pendukung dalam pengambilan keputusan medis berdasarkan data historis.

**Kata kunci: Prediksi, Penyakit Jantung, Naive Bayes**

### **ABSTRACT:**

*Background: The development of science has penetrated various fields, including health, where science is used for data processing. One example is the processing of heart disease patient data. However, most of the patient data has not been utilized optimally, so the information produced is less effective in supporting decision-making related to patient health, both those with and without heart disease. Decision-making based on accurate data or information can produce more precise predictions regarding the patient's health condition. Research Objectives: Based on these problems, this study aims to develop a heart disease prediction system using historical data. This system is designed to help the decision-making process more accurately by utilizing the Naive Bayes algorithm and the Python programming language. Research Methods: This study uses a quantitative method with a descriptive approach. The Naive Bayes algorithm is applied to analyze heart disease datasets to produce accurate predictions. Research Results: The results of the study show that the prediction system built is able to identify heart disease with a fairly high level of accuracy, namely 82%, 83%, and 84%. Testing was carried out using a dataset consisting of 1,024 patient data. Conclusion: It can be concluded that the prediction system based on the Naive Bayes algorithm is effective in predicting heart disease with good accuracy. This system can be a supporting tool in medical decision making based on historical data.*

*Keyword: Prediction, Heart Disease, Naive Bayes*

## PENDAHULUAN

Perkembangan ilmu pengetahuan saat ini telah mencakup berbagai disiplin ilmu, termasuk pemanfaatan teknologi untuk pengolahan data di bidang kesehatan. Salah satu penerapannya adalah pada pengelolaan data pasien penyakit jantung. Namun, sebagian besar data pasien belum diolah secara optimal, sehingga informasi yang dihasilkan kurang efektif untuk mendukung pengambilan keputusan klinis, baik bagi pasien yang terdiagnosis penyakit jantung maupun yang belum.

Dalam praktik medis, prediksi dan diagnosis sering kali didasarkan pada intuisi dokter daripada analisis data historis. Pendekatan ini berisiko menyebabkan ketidakakuratan dalam penentuan tindakan medis. Padahal, jantung merupakan organ vital yang berfungsi memompa darah ke seluruh tubuh melalui pembuluh darah. Istilah "penyakit jantung" secara umum merujuk pada berbagai gangguan fungsi jantung, termasuk penyakit jantung koroner, kelainan bawaan, dan gagal jantung.

Seiring kemajuan teknologi, terdapat peluang untuk mengembangkan sistem prediksi berbasis data guna meningkatkan akurasi diagnosis. Parameter seperti tekanan darah, jenis nyeri dada, kadar kolesterol, dan faktor risiko lainnya dapat dijadikan dasar dalam membangun model prediktif. Salah satu algoritma yang potensial untuk diterapkan adalah Naïve Bayes, yang dikenal efisien dalam menangani dataset besar serta relatif mudah diimplementasikan.

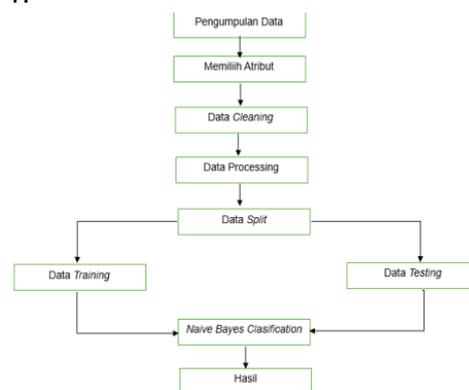
Penelitian ini bertujuan untuk membandingkan kinerja algoritma Naïve Bayes dengan pendekatan

sebelumnya menggunakan dataset penyakit jantung dari Kaggle.com yang terdiri dari 1.024 entri. Dengan mengembangkan sistem berbasis pemrograman, diharapkan dapat dihasilkan prediksi yang lebih akurat dalam menentukan risiko penyakit jantung pada pasien.

## METODE

Penelitian ini merupakan penelitian terapan (applied research) berbasis komputasional dengan pendekatan kuantitatif-eksperimental, yang bertujuan untuk membangun dan menguji sistem deteksi penyakit jantung menggunakan algoritma Naive Bayes. Penelitian bersifat interdisipliner, menggabungkan bidang ilmu komputer (khususnya deep learning dan computer vision) dengan ilmu medis.

Penelitian ini mengembangkan pendekatan pembelajaran mesin untuk deteksi penyakit jantung menggunakan algoritma Naive Bayes. Diagram alir penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

Proses penelitian diawali dengan pengumpulan dataset yang diperoleh dari situs web Kaggle.com. Dataset ini mencakup 14 atribut klinis pasien penyakit jantung yang meliputi jenis kelamin (sex), usia (age), tekanan

darah saat istirahat (trestbps), tipe nyeri dada (cp), kadar gula darah puasa (fbs), tingkat kolesterol (chol), hasil elektrokardiografi istirahat (restecg), detak jantung maksimum (thalach), angina akibat olahraga (exang), depresi ST akibat olahraga (oldpeak), kemiringan segmen ST (slope), jumlah pembuluh darah utama (ca), hasil scan thallium (thal), serta kondisi target (condition) yang menunjukkan ada atau tidaknya penyakit jantung. Variabel condition ini berperan sebagai atribut dependen dalam penelitian ini. Tahap selanjutnya adalah proses seleksi atribut, Proses pembersihan data dilakukan untuk menghilangkan noise dan memastikan kualitas data dengan menangani nilai-nilai yang hilang maupun outlier.

Klasifikasi menggunakan algoritma Naive Bayes diterapkan dengan menghitung probabilitas berdasarkan teorema Bayes yang mengasumsikan independensi antar fitur. Seluruh proses perhitungan diimplementasikan secara otomatis melalui sistem komputasi yang dikembangkan, bukan secara manual. Alur pengolahan data dimulai dari persiapan data training, pemisahan data testing, perhitungan probabilitas untuk data testing, implementasi sistem otomatis algoritma Naive Bayes, pemrosesan non-manual melalui sistem, hingga evaluasi hasil dan metrik akurasi untuk setiap skenario pengujian yang dilakukan. Pendekatan sistematis ini memungkinkan pemerolehan hasil yang konsisten dan dapat dipertanggungjawabkan secara ilmiah.

## **HASIL DAN PEMBAHASAN**

### **A. HASIL**

Berdasarkan pengujian yang dilakukan terhadap tiga variasi

pembagian dataset, diperoleh hasil metrik evaluasi yang komprehensif. Pada data testing 20%, confusion matrix menunjukkan nilai True Positive (TP) sebesar 85 dan True Negative (TN) sebesar 87, dengan False Positive (FP) 19 dan False Negative (FN) 14. Perhitungan akurasi mencapai 0.83 (83%), sementara sensitivitas (recall) sebesar 0.85 dan presisi 0.81. Nilai F1-Score yang merupakan harmonic mean antara recall dan presisi berada pada angka 0.82.

Untuk data testing 30%, performa model sedikit lebih baik dengan akurasi 0.84 (84%), dimana TP mencapai 128 dan TN 131. Presisi dan recall masing-masing 0.83 dan 0.84 dengan F1-Score konsisten di 0.83. Adapun pada data testing 40%, model menunjukkan akurasi 0.81 (81%) dengan TP 166 dan TN 169. Meskipun recall tetap tinggi (0.83), presisi mengalami penurunan menjadi 0.79 yang berdampak pada F1-Score 0.80. Secara komparatif, seluruh metrik tertinggi dicapai pada pembagian data 30% dengan akurasi 84%, presisi 83%, recall 84%, dan F1-Score 83%.

### **B. PEMBAHASAN**

Hasil penelitian mengungkapkan bahwa pembagian data 70:30 (training:testing) menghasilkan performa optimal dibandingkan variasi lainnya. Tingginya nilai recall pada semua skenario (0.83-0.85) menunjukkan kemampuan model yang baik dalam mengidentifikasi pasien positif penyakit jantung. Namun, penurunan presisi pada data testing 40% (0.79) mengindikasikan adanya peningkatan false positive, yang dalam konteks medis dapat berarti overdiagnosis. Nilai F1-Score yang stabil di atas 0.80 pada semua

pengujian menegaskan keseimbangan antara presisi dan recall.

Keberhasilan model Naive Bayes dalam penelitian ini didukung oleh karakteristik dataset yang memenuhi asumsi independensi fitur, serta distribusi data yang relatif seimbang antara kelas positif dan negatif. Hasil akurasi 84% pada data testing 30% membuktikan efektivitas algoritma ini untuk klasifikasi penyakit jantung berbasis parameter klinis.

### **KESIMPULAN**

Penelitian ini berhasil membuktikan bahwa model Naive Bayes mampu memprediksi penyakit jantung dengan akurasi hingga 84% menggunakan pembagian data training-testing 70:30. Metrik evaluasi yang komprehensif (akurasi, presisi, recall, dan F1-Score) secara konsisten menunjukkan performa terbaik pada skenario ini. Tingginya nilai recall (0.84) menjadikan model ini andal untuk skrining awal pasien berisiko, sementara presisi yang memadai (0.83) meminimalkan kesalahan diagnosis.

Kestabilan performa model dibuktikan melalui pengujian tiga skenario berbeda, dengan selisih akurasi hanya 1-3% antar variasi pembagian data. Hasil ini memperkuat posisi Naive Bayes sebagai algoritma yang efektif untuk klasifikasi data medis, khususnya pada dataset dengan karakteristik fitur independen dan ukuran sampel moderat (1024 instance).

### **DAFTAR PUSTAKA**

Darmajaya Kusri, dan Luthfi, E.M. 2009. *Algoritma Data Mining*. Yogyakarta : CV. Andi Offset

Fatta, Al Hanif. 2007. *Analisis & Perancangan Sistem Informasi*

untuk Keunggulan Bersaing Perusahaan & Organisasi Modern. Yogyakarta : CV. Andi Offset

Fatta, Al Hanif. 2007. *Analisis & Perancangan Sistem Informasi untuk Keunggulan Bersaing Perusahaan & Organisasi Modern*. Yogyakarta : CV. Andi Offset

Hasanah, Q., Andrianto, A. dan Hidayat, M.A. 2018. *Sistem Informasi Posyandu Ibu Hamil dengan Penerapan Klasifikasi Resiko Kehamilan Menggunakan Metode Naive Bayes*

Kusuma, P.J., 2013. *Data Mining Untuk Mendiagnosa Penyakit Infeksi Saluran Pernafasan (ISPA) Menggunakan Metode Naive Bayes*

Mustakim, 2013. *Implementasi Data Mining Untuk Identifikasi Pola Penyakit Ispa Dengan Menggunakan Algoritma Apriori (Studi Kasus Di Uptd Puskesmas Bae Kabupaten Kudus)*

Nurjoko, dan Kurniawan, H. 2016. *Aplikasi Datamining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Apriori di IIB Darmajaya Bandar Lampung. Informatics and Bussiness Institute*

Nurjoko, dan Kurniawan, H. 2016. *Aplikasi Datamining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Apriori di IIB Darmajaya Bandar Lampung. Informatics and Bussiness Institute*

Priyanti, E., 2017. *Penerapan Algoritma Naive Bayes Untuk Klasifikasi Bakteri Gram-Negatif*, Vol. 3, No.2

Saputra, R.A., 2014. Komparasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Penyakit Tuberculosis (Tb): Studi Kasus Puskesmas Karawang Sukabumi

Septiani dan Rijati, 2013. Klasifikasi Kehamilan Risiko Tinggi Pada Ibu Hamil Puskesmas Wonotunggal Batang Menggunakan Metode Naive Bayes