

**METODE NAÏVE BAYES DALAM MACHINE LEARNING  
UNTUK MEMPREDIKSI PENYAKIT JANTUNG DALAM TUBUH**

**Diana Lestari<sup>1</sup>**

<sup>1,2</sup>Sistem Informasi, STMIK Royal Kisaran

Email: <sup>1</sup>dianadiana@gmail.com

**Abstract:** The death rate due to heart disease is also a big record in Indonesia. The number of people with heart disease in Indonesia continues to increase, compared to developed countries in the European continent. Currently, many people with heart disease also come from the lower middle class in terms of the economy due to the lack of knowledge about maintaining heart health and the high cost of routine checks so that people ignore heart health. The purpose of this research is to apply the Naïve Bayes method to know how to predict someone with heart disease and to test the application of machine learning in predicting heart disease and to prove that the Naïve Bayes method can provide accurate results. the Naïve Bayes Algorithm method, produces accuracy with an average accuracy of above 90%.

**Keywords:** naïve bayes; predictions; heart disease.

**Abstrak:** Tingkat kematian karena penyakit jantung juga menjadi catatan besar di Indonesia. Jumlah penderita penyakit jantung di Indonesia terus meningkat, dibandingkan dengan negara maju di benua eropa. Saat ini penderita penyakit jantung juga banyak dari masyarakat kalangan menengah ke bawah dalam segi ekonomi dikarenakan minimnya pengetahuana untuk menjaga kesehatan jantung dan mahalnya biaya pemeriksaan rutin sehingga masyarakat mengabaikan kesehatan jantung. Tujuan diadakannya penelitian ini Menerapkan metode Naïve Bayes untuk mengetahui memprediksi seseorang terkena penyakit jantung dan Menguji penerapan *machine learning* dalam memprediksi penyakit jantung dan membuktikan metode metode Naïve Bayes dapat memberikan hasil akurasi yang akurat. metode Algoritma Naïve Bayes, menghasilkan akurasi dengan rata-rata akurasi diatas 90%.

**Kata Kunci:** naïve bayes; prediksi; penyakit jantung.

## **PENDAHULUAN**

Perkembangan dan pertumbuhan teknologi informasi yang berkembang pesat di dunia ini memerlukan kecerdasan buatan untuk mengolah berbagai data informasi. Berbagai teknologi telah dikembangkan oleh manusia untuk memenuhi keinginan manusia dengan tujuan bahwa kita tidak perlu lagi mencoba melakukan latihan yang melelahkan atau untuk mendapatkan hasil yang paling akurat. perkembangan teknologi dunia yang begitu pesat juga mempengaruhi kemajuan teknologi pada kesehatan manusia contohnya seperti kesehatan pada Jantung.

Jantung merupakan organ tubuh yang memiliki fungsi penting dalam tubuh manusia untuk menyediakan oksigen sebagaimana tugas jantung juga sebagai pusat yang memompa darah ke seluruh tubuh melibatkan pembuluh darah. Terlepas dari fungsi pentingnya jantung juga merupakan organ yang sangat rentan diserang penyakit akibat menumpuknya lemak dalam sel darah yang menyebabkan penyumbatan dalam sistem pembuluh darah hingga dapat menyebabkan penyakit jantung berbahaya lainnya[1]. Penyakit jantung yang terjadi di Indonesia memiliki kondisi yang serius dimana penyakit ini tidak melihat usia dari penderitanya. Berdasarkan SRS (Survei Sample Registration System) sebesar 12,9% penyebab kematian penduduk indonesia disebabkan oleh penyakit jantung.

Tingkat kematian karena penyakit jantung menjadi suatu catatan besar dimana mengingat bahwa penderita jantung di indonesia lebih banyak dari negara maju di benua eropa. Pasien penderita penyakit jantung banyak yang berasal dari keluarga yang menengah kebawah kehidupan ekonominya. Hal tersebut disebabkan oleh mininmnya kesadaran untuk menjaga kesehatan jantung dan mahalnya biaya yang harus dikeluarkan masyarakat untuk melakukan pemeriksaan rutin serta membutuhkan waktu yang lama untuk mendapatkan hasil diagnosa penyakit jantung atau tidak karena akan melalui proses pemeriksaan laboratorium yang

dilakukan oleh dokter dan untuk pemeriksaan ini dilakukan secara bertahap[1]. Maka demikian dengan adanya permasalahan diatas penelitian ini bertujuan untuk memprediksi penyakit jantung dalam *machine learning* agar dapat menyajikan informasi diagnosa penyakit jantung akurat dengan menambahkan kuantitas data yang diuji sehingga dapat menghasilkan sebuah informasi yang tepat. Untuk penelitian ini peneliti akan menggunakan pengambilan data pada data sekunder penyakit jantung yang didapat dari website Kaggle. Penelitian ini akan melakukan pengolahan data menggunakan *machine learning* dengan metode Naïve Bayes, Hasil guna memperoleh tingkat akurasi pada dataset yang akan digunakan.

Algoritma Naïve Bayes adalah sebuah metode klasifikasi dengan teorema bayes pada ilmu statistik [2]. Naïve Bayes dapat digunakan untuk memprediksi kemungkinan [3]. Penerapan penelitian dalam bidang Kesehatan, dengan menggunakan metode Naïve Bayes yang telah banyak digunakan salah satunya yang dilakukan oleh [4] tentang “implementasi *machine learning* untuk prediksi penyakit ginjal”, dalam penelitian tersebut metode Naïve Bayes memiliki kinerja dengan tingkat akurasi prediksi rata-rata akurasi sebesar 87.27% .

## **METODE**

### **1. Peralatan Riset**

Perangkat yang akan digunakan dalam penelitian ini berupa komputer 64 bit dengan RAM 4GB, *software anaconda navigator*, *visual studio code* dan HTML untuk membangun *deploy*.

### **2. Data Riset**

Data yang digunakan terdapat 5 variabel diantaranya *Age* (usia), *Sex* (jenis kelamin), *CP/Chest Pain Type* (jenis nyeri dada), *trtbps/Trestbps* (tekanan darah), *Chol* (kolesterol) dan (*Heart Disease*) sebagai kolom target/label.

**Tabel 1.** Dataset Penyakit Jantung

Age	Sex	ChestPainType	RestingBP	Cholesterol	HeartDisease
40	M	ATA	140	289	0
49	F	NAP	160	180	1
37	M	ATA	130	283	0
48	F	ASY	138	214	1
54	M	NAP	150	195	0
39	M	NAP	120	339	0

## HASIL DAN PEMBAHASAN

### 3. Prediksi dengan Model Naive Bayes

Naïve bayes digunakan untuk memprediksi sebuah kemungkinan dengan pola data yang sudah ada sebelumnya. Naïve bayes sering digunakan dalam penelitian terdahulu untuk memprediksi sebuah studi kasus dan menghasilkan akurasi yang baik seperti pada penelitian [5] dimana dalam penelitian tersebut Naïve bayes memiliki akurasi diatas 60% sehingga dapat dikategorikan sebagai *good clasification*. Pada penelitian ini prediksi menggunakan Naïve Bayes menggunakan persamaan sebagai berikut [6]:

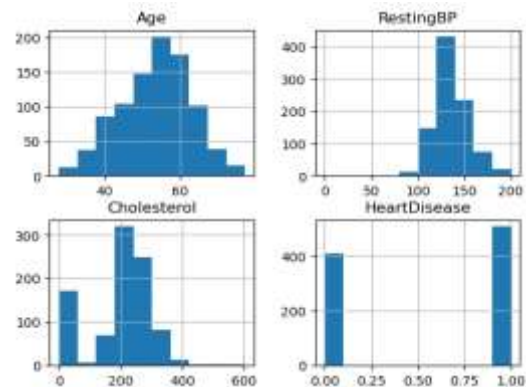
$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (1)$$

Deskripsi :

X : Data Class  
H : Class Spesific  
P(H|X) : Probabilitas hipotesa H didasarkan pada keadaan X  
P(H) : Probabilitas hipotesa H  
P(X|H) : Probabilitas X didasarkan pada keadaan H  
P(X) : Probabilitas hipotesa X

### 1. Visualisasi Data

Visualisasi data disajikan agar penerima informasi lebih mudah memahami data infrmasi yang disajikan. *Scatter* plot menggunakan *library* yaitu “*matplotlib.pyplot*”. Dalam *machine learning*, *matplotlib.pyplot* merupakan sebuah modul atau fungsi yang digunakan untuk menampilkan gambar, grafik dan bentuk visualisasi lainnya dalam bahasa python. Pada penelitian ini visualisasi data menggunakan *scatter* plot yang bertujuan untuk melihat plot sebaran data dan mengamati. Berikut ini adalah visualisasi data grafik pada penyakit jantung.



**Gambar 1.** Grafik Pada Penyakit Jantung

Pada gambar tersebut, terdapat 4 grafik terjadinya penyakit jantung yaitu : *Age* (usia), *Sex* (jenis kelamin), *CP/Chest Pain Type* (jenis nyeri dada), *trtbps/Trestbps* (tekanan darah), *Chol* (kolesterol). Berdasarkan grafik diatas, dapat diketahui bahwa banyaknya gejala yang mengakibatkan penyakit jantung adalah BP (radang paru-paru).

## 2. Pengujian Data dengan Naive Bayes

Pengujian penelitian ini dilakukan secara berulang dengan menggunakan 4 teknik pembagian data rasio dalam setiap pengujiannya, hal tersebut memiliki tujuan untuk mengetahui hasil pengujian yang paling efektif dari 4 pembagian data. Hasil perolehan dari pengujian data menggunakan Naive Bayes disajikan dalam bentuk tabel sebagai berikut:

**Tabel 1.** Pengujian Data

Nomor	Pembagian Data Rasio	Hasil
1	90 : 20	86%
2	80 : 30	91%
3	70 : 40	76%

Pada tabel 1 merupakan hasil dari pengujian dengan 4 metode pembagian data. Dapat dilihat bahwa prediksi penyakit jantung dengan metode Naive Bayes memiliki akurasi paling akurat pada pembagian data rasio 80 : 30 yang artinya 80% data latih dan 30% data uji dan hasil yang dikeluarkan tergolong dalam *good classifier* sebesar 90%. Untuk menguji hasil akurasi maka dilakukan dengan *confusion matrix* sebagai berikut ini:



**Gambar 2.** Visualisasi *Confusion Matrix*

Pada gambar tersebut diperoleh dari hasil :  
 Nilai benar :  $27 + 9 = 36$   
 Nilai salah :  $39 + 6 = 45$   
 Hasilnya :  $36/45 = 0,8$

Adapun nilai akurasinya terlihat pada gambar dibawah ini :

	precision	recall	f1-score	support
0	0.90	0.95	0.92	37
1	0.96	0.93	0.94	55
accuracy			0.93	92
macro avg	0.93	0.94	0.93	92
weighted avg	0.94	0.93	0.94	92

**Gambar 3.** Nilai Akurasi

Berdasarkan perolehan *confusion matrix* dapat diketahui bahwa *Naive Bayes* menghasilkan nilai diatas 90% di masing-masing fungsi *confusion matrix* dimana perolehan nilai antara lain 90% penderita jantung pada *precision*, 93% pada *recall*, dan 95% pada *support*. Maka dengan demikian dapat dikatakan bahwa metode *Naive Bayes* tepat untuk diimplementasikan dalam prediksi klasifikasi penyakit jantung pada tubuh manusia.

## KESIMPULAN

Pada penelitian diperoleh kesimpulan dari hasil proses perhitungan dengan metode Naive Bayes dalam *machine learning* bahwa metode Naive Bayes merupakan metode yang tepat untuk memprediksi penyakit jantung dan akurat apabila diimplementasikan dengan deploy sehingga pengguna dapat mengoperasikan aplikasi dan memahami fitur-fitur aplikasi untuk memprediksi kondisi kesehatan jantung secara mandiri dan diharapkan dapat membantu untuk mengatasi permasalahan rendahnya informasi masyarakat akan kesehatan jantung yang disebabkan oleh mahalnya biaya pemeriksa rutin.

**DAFTAR PUSTAKA**

- [1] Aswir and H. Misbah, *Photosynthetica*, vol. 2, no. 1, pp. 1–13, 2018, [Online]. Available: <http://link.springer.com/10.1007/978-3-319-76887-8>  
<http://link.springer.com/10.1007/978-3-319-93594-2>  
<http://dx.doi.org/10.1016/B978-0-12-409517-5.00007-3>  
<http://dx.doi.org/10.1016/j.jff.2015.06.018>  
<http://dx.doi.org/10.1038/s41559-019-0877-3>
- [2] Septiani, W. D. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4. 5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis. *Jurnal Pilar Nusa Mandiri*, 13(1), 76-84.
- [3] Apriliah, W., Kurniawan, I., Baydhowi, M., & Haryati, T. (2021). Prediksi kemungkinan diabetes pada tahap awal menggunakan algoritma klasifikasi Random Forest. *Sistemasi: Jurnal Sistem Informasi*, 10(1), 163-171.
- [4] Alotaibi, F. S. (2019). Implementation of machine learning model to predict heart failure disease. *International Journal of Advanced Computer Science and Applications*, 10(6).
- [5] Saputro, I. W., & Sari, B. W. (2020). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal*, 6(1), 1-11.
- [6] M. A. Bianto, K. Kusriani, and S. Sudarmawan, “Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes,” *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 75, 2020, doi: 10.24076/citec.2019v6i1.231.

