

Implementasi Metode Naive Bayes Untuk Memprediksi Resiko Penyakit Jantung

Muhammad Ranjasmara¹⁾, Amal Khairan²⁾, Firman Tempola³⁾ Rosihan⁴⁾

^{1,2,3,4} Fakultas Teknik, Universitas Khairun,

¹pti07351511117@gmail.com, ²Amal.khairan@unkhair.ac.id,

³firman.tempola@unkhair.ac.id, ⁴rosihan@unkhair.ac.id

Abstrak

Kematian akibat penyakit jantung terus meningkat dan tak mengenal usia muda dan tua. *World Health Organization* menyebutkan 7,3 juta penduduk dunia meninggal akibat dari penyakit jantung. Bahkan disebutkan penyakit jantung adalah salah penyakit nomor satu paling mematikan. Untuk itu penting diketahui resiko dari penyakit jantung dengan menerapkan model-model yang ada pada *machine learning*. Tujuan dari penelitian ini yaitu untuk mengimplementasikan metode Naive Bayes untuk memprediksi penyakit jantung, serta dilakukan uji kinerja algoritma dengan menghitung presisi, recall dan akurasi. Adapun Kriteria-kriteria yang digunakan pada penelitian ini yaitu umur, jenis kelamin, jenis sakit dada, tekanan darah, kolesterol, kadar gula, elektrokardiografi, tekanan jantung, angina induksi, old-peak, segmen_st, Fluoroskopi, denyut jantung. Sedangkan class yang diprediksi ada 2 beresiko dan tidak beresiko. Hasil dalam penelitian ini menunjukkan bahwa metode berhasil memprediksi atau mengklasifikasi pasien beresiko penyakit jantung dan tidak beresiko penyakit jantung dengan persentase precision 90%, recall 100% serta mendapatkan akurasi 92.85% dan termasuk excellent classification.

Kata kunci: Resiko penyakit jantung, machine learning, naive bayes

Abstract

Deaths from heart disease continue to rise and do not recognize young and old. The World Health Organization says 7.3 million people have died from heart disease. It even mentioned heart disease is one of the deadliest number one diseases. Therefore, it is important to know the risk of heart disease by applying existing models to machine learning. The purpose of this study is to implement the Naive Bayes method to predict heart disease, as well as test the performance of algorithms by calculating precision, recall and accuracy. The criteria used in this study are age, gender, type of chest pain, blood pressure, cholesterol, sugar levels, electrocardiography, heart pressure, angina induction, old-peak, segmen_st, fluoroscopy, heart rate. Meanwhile, the class predicted there were 2 at risk and not at risk. The results in this study showed that the method successfully predicted or classified patients at risk of heart disease and not at risk of heart disease with a percentage precision of 90%, recall 100% and obtained 92.85% accuracy and advanced excellent classification.

Keywords: Risk of Heart disease, Machine learning, Naive bayes

PENDAHULUAN

Jantung merupakan organ manusia paling vital yang berperan dalam sistem peredaran darah. Penyakit jantung adalah sebuah kondisi dimana jantung tidak dapat melaksanakan tugasnya dengan baik. Penyakit jantung juga tidak menyerang hanya pada usia tua melainkan beberapa kasus kematian akibat penyakit jantung juga pada usia muda.

Data WHO (*World Health Organization*) menyatakan bahwa sebanyak 7,3 juta penduduk dunia meninggal dikarenakan penyakit jantung. Meskipun penyakit jantung merupakan penyakit yang tidak menular, penyakit ini merupakan jenis penyakit yang mematikan nomor satu di dunia [1].

Oleh karena itu perlu adanya aplikasi yang bisa memprediksi resiko penyakit jantung,

karena dengan memprediksi gejala-gejala yang di alami dapat menanggulangi resiko yang lebih parah.

Naive Bayes merupakan metode probabilistic pengklasifikasian sederhana berdasarkan Teorema Bayes dimana pengklasifikasian dilakukan melalui training set sejumlah data secara efisien. Naive bayes mengasumsikan bahwa nilai dari sebuah input atribut pada kelas yang diberikan tidak tergantung dengan nilai atribut yang lain, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Penelitian dengan menerapkan algoritma naïve bayes pernah dilakukan oleh beberapa peneliti sebelumnya [2] pada penentuan UKT. [3] pada penentuan status gunung berapi. Sistem ini juga dilakukan pengujian kinerja sistem sebagaimana pernah dilakukan oleh [4],[6].

METODE PENELITIAN

Penelitian menerapkan metode machine learning dalam hal ini adalah naïve bayes classifier dengan dataset yang digunakan dari ucl machine learning. Sedangkan sistem yang dibangun menggunakan Bahasa pemograman PHP MYSQL. Sistem ini juga dilakukan uji kinerja sistem dengan menggunakan uji akurasi, presisi dan recall.

A. Machine learning

Machine Learning didefinisikan sebagai sebuah tipe kecerdasan buatan (AI) yang menyediakan komputer dengan kemampuan untuk belajar dari data, tanpa secara eksplisit harus mengikuti instruksi yang terprogram. Definisi lain Machine Learning (ML) yaitu pembelajaran mesin merupakan pendekatan dalam AI yang banyak digunakan untuk menggantikan atau menirukan perilaku manusia untuk menyelesaikan masalah atau melakukan otomatisasi. Sesuai namanya, Machine Learning mencoba menirukan bagaimana proses manusia atau makhluk cerdas belajar dan menggeneralisasi. Setidaknya ada dua aplikasi utama dalam Machine Learning yaitu, klasifikasi dan prediksi [5].

B. Naïve bayes classifier

Naive Bayes merupakan metode probabilistic pengklasifikasian sederhana berdasarkan Teorema Bayes dimana pengklasifikasian dilakukan melalui training set sejumlah data secara efisien. Naive bayes mengasumsikan bahwa nilai dari sebuah input atribut pada kelas yang diberikan tidak tergantung dengan nilai atribut yang lain. Teorema Bayes sendiri dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes [6]. Untuk memprediksi resiko penyakit jantung dengan naïve bayes memiliki langkah-langkah berikut.

Langkah pertama: menghitung mean dan standar deviasi setiap data numerik.

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \text{ atau } \bar{x} = \frac{x_1+x_2+x_3+\dots+x_n}{n} \dots\dots\dots(1)$$

Langkah kedua: Menghitung probabilitas setiap data diskrit.

$$P(E) = \frac{X}{N} \dots\dots\dots(2)$$

Langkah ketiga: Menghitung densitas gauss

$$f(x_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x-\mu}{2\sigma^2}} \dots\dots\dots(3)$$

Langkah keempat: Menghitung likelihood

$$L_x = f_{x_1} \times f_{x_2} \times \dots \times f_{x_n} \dots\dots\dots(4)$$

Langkah kelima: Menghitung probabilitas likelihood

$$P_{L_i} = \frac{L_i}{\sum_{i=1}^n L_i} \dots\dots\dots(5)$$

C. Akurasi

Pengukuran akurasi dapat menggunakan berbagai cara salah satunya menggunakan recognition rate [5].

$$Recognition\ Rate = \frac{\sum_{corret}}{\sum_{sample}} \times 100\% \dots\dots\dots(6)$$

D. Laplacian Smoothing

Dalam sebuah dataset yang besar, pemilihan data training secara random akan menyebabkan kemungkinan adanya nilai nol dalam model probabilitas. Nilai

nol ini akan menyebabkan Naive Bayes Classifier tidak dapat mengklasifikasi sebuah data inputan. Oleh karena itu diperlukan suatu metode smoothing yang dapat menghindari adanya nilai nol dalam model probabilitas [7].

$$P(X_k|C) = \frac{P X_k C + 1}{P C + |V|} \dots\dots\dots(7)$$

E. Dataset

Data yang digunakan pada penelitian ini berasal dari UCI Machine Learning Repository Heart Disease Data Set, Jumlah data yang digunakan sebanyak 303 data, dengan rincian 194 data berkategori tidak beresiko dan 109 data berkategori beresiko. Sedangkan kriteria yang digunakan ada 13 kriteria yaitu umur, jenis kelamin, jenis sakit dada, tekanan darah, kolestrol, kadar gula, elektrokardiografi, tekanan jantung, angina induksi, old-peak, segment_st, Fluoroskopi dan denyut jantung.

HASIL DAN PEMBAHASAN

A. Prediksi resiko penyakit jantung
 untuk memprediksi resiko penyakit jantung dengan naïve bayes maka tahap awal adalah dikumpulkan resiko penyakit jantung. Kemudian dilihat jenis data. Pada dataset resiko penyakit jantung yang diterapkan yaitu ada dua jenis data diskrit dan numerik sehingga tahapan penyelesaian harus menghitung seluruh nilai mean dan standar deviasi dari kategori atau kriteria dengan jenis data numerik. Dengan diasumsikan data uji yaitu dengan nilai kriteria umur (53 tahun), jenis kelamin (wanita), jenis sakit dada (typ_angina), tekanan darah (145), kolestrol (100), kadar gula (T), Elektrokardiografi (normal), Tekanan jantung (178.0), angina induksi (yes), old-peak (1.5), segmen_st (down), fluoroskopi (1.1) denyut jantung (normal), dengan jumlah data yang ditraining yaitu 203 data. Pada Tabel 1 merupakan hasil hitung nilai Mean dan standar deviasi pada kriteria umur.

Tabel 1. Hasil hitung mean dan standar deviasi

	Tidak beresiko	Beresiko
Banyak data	131	72
Jumlah	6556	4345
Mean	50,046	60,361
Standar deviasi	8,382	6,312

Seluruh kriteria dengan jenis data numerik harus dihitung nilai mean dan standar deviasi. Selanjutnya dihitung probabilitas dari kriteria dengan jenis data diskrit. Pada Tabel 2 merupakan hasil hitung probabilitas untuk kriteria jenis kelamin.

	Tidak beresiko	Beresiko	Tidak beresiko	Beresiko
Laki-laki	84/131	59/72	0,641	0,819
Wanita	47/131	13/72	0,359	0,181
Jumlah	131/131	72/72	1	1

Seluruh data yang bertipe kategori atau diskrit harus dihitung nilai probabilitasnya. Selanjutnya dihitung densitas gaus dengan menggunakan Persamaan 3. Hasil hitung densitas gaus ditunjukkan pada Tabel 3.

	Tidak Beresiko	Beresiko
umur	0.04474068008706	0.032028192676251
Tekanan darah	0.021101871784454	0.020386420849965
Kolestrol	0.00012026336158039	6.1392446835068E-10
Tekanan Jantung	0.020954516895628	0.0036889295692124
Old-Peak	0.3244151929944	0.2765230883679
Fluoroskopi	1.0252792953798E-8	0.38533631911848

Selanjutnya dihitung likelihood setiap kategori klasifikasi sesuai dengan data uji seperti berikut:

Untuk mendapatkan nilai likelihood dapat menggunakan persamaan 5.

$$\begin{aligned}
 \text{a. Likelihood Tidak Beresiko} &= 0.04474068008706 \times 0.021101871784454 \times \\
 &0.00012026336158039 \times 0.020954516895628 \times 0.3244151929944 \times \\
 &1.0252792953798E-8 = 7.91E-05
 \end{aligned}$$

$$\begin{aligned}
 \text{b. Likelihood Beresiko} &= 0.032028192676251 \times 0.020386420849965 \\
 &\times 6.1392446835068E-10 \times 0.0036889295692124 \times 0.2765230883679 \times \\
 &0.38533631911848 = 1.58E-03
 \end{aligned}$$

berdasarkan nilai likelihood dari dua kelas antara beresiko dan tidak beresiko pada data dari pasien diprediksi beresiko untuk penyakit jantung.

B. Implementasi sistem

Tahap awal ada dengan melakukan login pada sistem. Setelah berhasil melakukan login, admin di arahkan ke halaman utama sebagaimana ditunjukkan pada Gambar 1



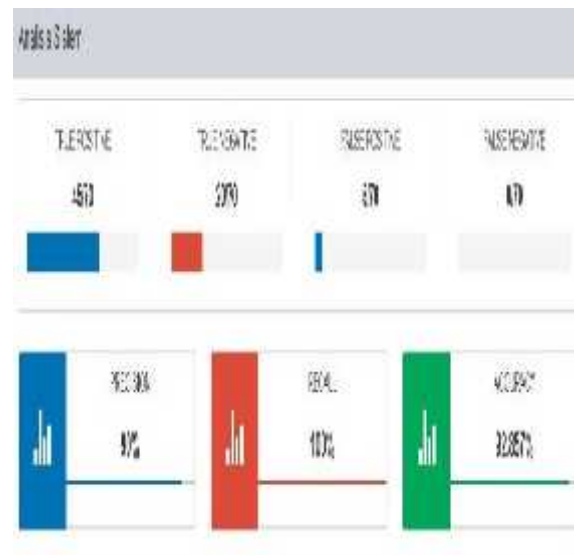
Gambar 1. Halaman Utama sistem

Pada halaman ini di lakukan prediksi terhadap data pasien, pengguna dapat meng-input data-data pasien. Adapun tampilannya sebagaimana ditunjukkan pada Gambar 2.

Gambar 2. Halaman Prediksi

C. Halaman Analisis sistem

Pada aplikasi ini juga dilengkapi dengan halaman untuk menganalisis kinerja dari algoritma. Dimana dilengkapi dengan uji confusion matrix yaitu presisi, recall dan akurasi. Halaman analisis sistem ditunjukkan pada Gambar 3.



Gambar 3. Halaman Analisis Sistem

Tabel 4. Perbandingan kinerja sistem

No.	Data latih	Data Uji	TP	TN	FP	FN	Akurasi
1	203	100	63	27	10	0	90%
2	100	203	126	55	18	4	89,16%
3	20	283	169	78	24	12	87%
4	70	70	45	20	5	0	92,85%

D. Kinerja algoritma naïve bayes

Pada tahap pengujian akurasi metode di lakukan untuk mencari nilai akurasi yang tertinggi dengan cara mengklasifikasikan kembali data uji pada dataset pasien penyakit jantung untuk mencari data benar dan data salah setelah di lakukan pembagian dataset dengan menggunakan metode hold out validation yaitu menentukan jumlah data latih yang akan digunakan pada dataset dan sisa dari dataset dijadikan sebagai data uji. Pengujian dilakukan sebanyak 4 kali pengujian dengan jumlah data latih dan data uji yang berbeda-beda. Adapun rumus yang digunakan untuk menghitung akurasi yaitu dengan persamaan 5. Adapun hasil pengujian ditunjukkan pada Tabel 4.

Berdasarkan hasil pengujian yang dilakukan, sistem aplikasi dengan metode naïve bayes mendapatkan hasil dengan akurasi tertinggi yaitu pada percobaan ke 1 dan ke 4 dengan 203 - 70 data latih dan 100 - 70 data uji, berhasil mendapatkan akurasi yaitu 90% dan 92.85%

KESIMPULAN

Menentukan kriteria-kriteria dari penelitian, dari kriteria itu di simpan kedalam *database*, data yang di simpan pada *database* akan digunakan untuk membagi data latih dan data uji setelah di tentukan data latih dan data ujinya dilakukanlah perhitungan menggunakan persamaan 1 untuk mencari nilai *mean* dan *standar deviasi*, setelah itu mencari nilai probabilitas data diskrit dengan persamaan 2, mencari nilai *densitas gauss* dengan persamaan 3 dari data *densitas gauss* dicari nilai *likelihood* menggunakan persamaan 4 dan nilai probabilitas *likelihood* persamaan 5. Dalam pengujian akurasi system dengan metode *naïve bayes* dilakukan sebanyak 103 kali pengujian, dimana hasil tertinggi itu

92.85% dengan data latih 70 data dan data uji 70 data sedangkan, akurasi terendah yaitu 87% dimana data latih 20 data dan data uji 283 data. Untuk penelitian selanjutnya dapat menggunakan data dari rumah sakit yang berada di indonesia dan dengan kasus penyakit jantung dapat menggunakan metode yang lain.

REFERENSI

- [1]. Lestari, M. E. I. (2014). Penerapan Algoritma Klasifikasi Nearest Neighbor (K-Nn) Untuk Mendeteksi Penyakit Jantung. 7(September 2010), 366-371.
- [2]. Tempola F, Muhammad M, and Khairan A, 2018 "Naive Bayes Classifier for Prediction of Volcanic Status in Indonesia," *Proc. - 2018 5th Int. Conf. Inf. Technol. Comput. Electr. Eng. ICITACEE 2018*, pp. 365-369
- [3]. Anas A, Khairan A dan Tempola F., 2019. Hybrid Fuzzy dan Naïve Bayes dalam penentuan Status UKT (Studi Kasus Program Studi Teknik Informatika Univeristas Khairun). *PROtek: Jurnal Ilmiah Teknik Elektro*. Vol. 6 (1). pp.6-11
- [4]. Baksir A H, Fuad A, Tempola F, Rosihan R, 2020. Prediksi Tingkat Kualitas Kesuburan Pria Dengan Jaringan Saraf Tiruan Backpropagation. *JIKO (Jurnal Informatika dan Komputer)*. Vol.3 (2). pp.107-112.
- [5]. Ahmad, A. (2017). Mengenal Artificial Intelligence , Machine Learning , Neural Network , Dan Deep Learning, (June).
- [6]. Dahri, D., Agus, F., & Khairina, D. M. (2016). Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Bidikmisi Universitas Mulawarman, 11(2).