Jurnal Teknik Elektro dan Komputasi (ELKOM)
Volume 6 Nomor 1 Maret 2024 ‖ Hal. 12-17
ISSN : 2685-1814 (*Print*)
ISSN : 2685-7677 (*Online*)

Klasifikasi Penyakit Hati Menggunakan Perbandingan Implementasi Algoritma *Naïve Bayes* Dan *K-Nearest Neighbor*

Valentino Simamora[1], Anita Desiani[1*], Irmeilyana[1]

Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sriwijaya
Jl. Raya Palembang – Prabumulih Km.32 Indralaya Indah, Kabupaten Ogan Ilir, Sumatera Selatan
E-mail: anita_desiani@unsri.ac.id

Naskah Masuk: 11 Juni 2023; Diterima: 03 Oktober 2023; Terbit: 31 Maret 2024

ABSTRAK

Abstrak **-** Hati merupakan organ kelenjar dalam tubuh manusia. Hati manusia memiliki bobot kira-kira mencapai 1200 hingga 1500 gram. Sebagai kelenjar terbesar dalam tubuh manusia, hati dapat terserang berbagai macam penyakit. Kita dapat melakukan klasifikasi mengenai penyakit hati yang bertujuan memperoleh jumlah rata-rata manusia yang terserang penyakit hati. Dengan penelitian ini, kita bisa membandingkan dan menyimpulkan algoritma mana yang paling tepat untuk diterapkan pada proses klasifikasi terhadap penyakit hati. Pada penelitian ini algoritma yang digunakan ialah algoritma pertama *Naïve Bayes* dan algoritma kedua *K-Nearest Neighbor* (K-NN). Dari hasil penelitian maka diperoleh bahwa *Naïve Bayes* memberikan nilai akurasi, presisi dan *recall* sebesar 85%-85,5% yang mana ini dapat dikatakan cukup baik namun belum baik. Sedangkan K-NN dapat memberikan nilai sempurna pada akurasi, presisi dan *recall* yaitu 100%. Maka algoritma yang terbaik dan dapat digunakan adalah algoritma K-NN.

Kata kunci: *Hati, Naïve Bayes, K-Nearest Neighbor, Perbandingan, Penyakit*

ABSTRACT

Abstract **-** *The liver is a glandular organ in the human body. The human heart weighs approximately 1200 to 1500 grams. As the largest gland in the human body, the liver can be attacked by various diseases. We can classify liver disease in order to obtain the average number of people with liver disease. With this research, we can compare and conclude which algorithm is the most appropriate to apply to the classification process for liver disease. In this study the algorithm used is the first Naïve Bayes algorithm and the second K-Nearest Neighbor (K-NN) algorithm. From the results of the study it was found that Naïve Bayes gave accuracy, precision and recall values of 85% -85.5% which can be said to be quite good but not good. While K-NN can give a perfect score on accuracy, precision and recall that is 100%. Then the best algorithm and can be used is the K-NN algorithm.*

Keywords: *Liver, Naïve Bayes, K-Nearest Neighbor, Comparison, Diseases*

1. PENDAHULUAN

Hati merupakan organ kelenjar di dalam tubuh manusia, sebagai kelenjaer terbesar, hati memiliki bobot yang mencapai sekitar 1200 hingga 1500 gram [1]. Hati berfungsi untuk menyaring (sebagai filter) setiap darah yang mengalir melalui vena porta, darah tersebut diterima dari usus kemudian hati akan mengubah dan menyimpan bahan-bahan makanan yang diterima [2]. Sebagai kelenjar terbesar, hati memiliki cara khusus untuk meregenerasi dirinya, yaitu dengan meningkatkan diferensial sel punca menjadi hepatosit atau kolangiosit, selain itu dapat juga dengan cara meningkatkan kecepatan mitosis hepatosit [3].

Hati dapat terserang berbagai jenis penyakit diantaranya kanker hati atau dengan nama lain Hepatoma, perlemakan hati non alkoholik, sirosis, abses hati, kolesistitis dan hepatitis [4]. Klasifikasi dapat digunakan untuk melakukan deteksi terhadap penyakit hati. Contoh klasifikasi yang dapat diterapkan adalah data mining.

ISSN : 2685-1814 (*Print*)
ISSN : 2685-7677 (*Online*)

Sebelumnya, Institut Adhi Tama Surabaya pernah melakukan pengklasifikasian mengenai penyakit hati [5]. Klasifikasi tersebut dilakukan dengan menggunakan algoritma *Support Vector Machines* (SVM), algortima K-NN dan algoritma *Naïve Bayes*. Hasil yang didapat menyatakan bahwa algoritma SVM memperoleh hasil hingga 84,62%, algortima *Naïve Bayes* memperoleh hasil hingga 82,42% dan algortima K-NN memperoleh hasil hingga 63,74-68,13%. Artinya pada penelitian tersebut menyatakan bahwa algortima yang paling cocok digunakan untuk dataset tersebut adalah algoritma SVM. Pada penelitian lainnya yang dilakukan oleh Rhyzoma Grannata Rafsanjani dengan metode *Naïve Bayes* dan *Certainty Factor* menyatakan bahwa tingkat akurasi yang diperoleh mencapai hingga 88% [6]. Artinya dalam penelitian tersebut menyatakan bahwa metode tersebut cukup baik untuk digunakan pada klasifikasi atau deteksi penyakit hati.

Pada dasarnya algoritma atau metode yang dapat digunakan sangat banyak, tergantung dengan kebutuhan dan kecocokan algoritma yang digunakan terhadap dataset yang dimiliki. Kita ambil contoh algoritma misalnya *Naïve Bayes*, kelebihan dari algoritma *Naïve Bayes* ini adalah dapat memberikan akurasi dengan kecepatan yang cukup tinggi ketika digunakan pada database yang besar dan hanya membutuhkan data training dengan jumlah yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Adapun kelemahan dari penggunaan algortima *Naïve Bayes* yaitu jika probabilitas kondisionalnya adalah nol mengakibatkan probabilitas prediksi akan bernilai nol juga, sehingga algoritma *Naïve Bayes* tidak berlaku. Selain *Naïve Bayes*, algortima lainnya seperti algoritma K-NN (*K-Nearest Neighbor*), Kelebihan algoritma K-NN ini ialah mudah dipelajari karena sederhana, proses pelatihan yang sangat cepat, tetap dapat digunakan pada data yang memiliki bias, dan tetap efektif untuk digunakan pada data berjumlah besar. Kekurangan dari algoritma K-NN adalah keterbatasan memori, biasnya nilai k, dipengaruhi oleh data-data yang tidak relevan dan komputasi kompleks.

Penelitian kali ini menggunakan 2 algoritma yaitu *Naïve Bayes* dan algoritma K-NN untuk melakukan klasifikasian pada dataset mengenai seseorang sedang terserang penyakit hati atau tidak. Kita akan menghitung nilai-nilai presisi, *recall* serta akurasi dari dataset tersebut. Kedua algoritma yaitu algoritma *Naïve Bayes* dan algortima K-NN akan digunakan dan setiap hasilnya akan dibandingkan untuk memperoleh kesimpulan mengenai algoritma yang paling cocok untuk dataset tersebut ketika melakukan klasifikasi.

2. METODE PENELITIAN

Penelitian ini menggunakan dua algoritma yaitu *Naïve Bayes* dan K-NN. Pertama cari dan tentukan dataset mengenai penyakit hati.

2.1. Deskripsi Data

Dataset yang akan kita gunakan untuk melakukan penelitian ini diambil dari situs Kaggle (https://www.kaggle.com/datasets/ johnsmith88/heart-disease-dataset) dalam bentuk file csv. Ada 14 atribut diantaranya 13 atribut sebagai ciri dan sisa satu sebagai atribut target. Target juga dibagi menjadi 2 label, yaitu pertama 0 = tidak ada penyakit dan yang kedua 1 = ada penyakit. Didalam dataset ini memiliki 1025 data. Atribut yang digunakan pada dataset ini yaitu *age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, target.*

Tabel 1. Penjelasan dan Arti dri Atribut

Attribute	***Feature Meaning***
Target/Target	0 = tidak ada penyakit, 1 = ada penyakit
Age/Umur	
Sex/Jenis Kelamin	0 = female, 1 = male
Chest Pain/ NyeriDada	0 = tidak nyeri, 1 = nyeri ringan, 2 = nyeri, 3 = sangat nyeri
Trestbps/ Tekanan DarahRendah	
Cholestrol/Kolestrol	
Fbs / Gula DarahNormal	0 = salah, 1 = benar
Restecg/ Elektrokardiografi	0 = tidak sehat, 1 = cukup sehat, 2 = sehat
Thalach/ DetakJantung Maksimum Tercapai	
Exang/ KestabilanInduksi Angina	0 = tidak stabil, 1 = stabil
Oldpeak / DepresiST yang diinduksioleh olahraga relatif terhadapistirahat	

Attribute	*Feature Meaning*
Slope/ Kemiringansegmen	
Ca / nomor pembuluh darahutama	
Thal	0 = normal, 1 = cacat tetap, 2 = cacat reversibel

2.2. *Processing* Data

Pada tahap ini kita akan melakukan beberapa hal seperti membuang data yang terduplikasi, melakukan pemeriksaan terhadap data yang tidak konsisten, memperbaiki setiap kesalahan yang ada didalam data, salah satu contohnya kesalahan cetak [6]. Didalam dataset ini tidak terdapat atribut yang harus dibuang atau dihapus, dengan begitu dapat dinayatakan bahwa seluruh atribut yaitu sebanyak 14 atribut akan digunakan dalam klasifikasi pada penelitian ini.

Kita akan menggunakan teknik *presentase split* dengan perbandingan 8:2. Kita membutuhkan data yang digunakan dalam proses pembelajaran atau latih yaitu data *training.* Selain itu kita juga membutuhkan data yang akan diuji yaitu data *testing.* Dengan menggunakan perbandingan 8:2 maka artinya sebanyak 80% dari data akan dijadikan sebagai data *training* (data latih) dan sebanyak 20% dari data akan dijadikan data *testing* (data uji).

2.3. Algoritma *Naïve Bayes*

Naïve Bayes diperkenalkan oleh Thomas Bayes yang merupakan seorang ilmuwan dari inggris, ini merupakan pengklasifikasian dengan metode statistik dan metode probabilitas [7]. Berikut ini merupakan langkah-langkah melakukan klasifikasi dengan menggunakan algoritma *Naïve Bayes*:

a. Melakukan perhitungan terhadap jumlah kategori pada setiap variable yang ada
b. Melakukan perhitungan terhadap peluang dari setiap kategori
c. Menentukan jumlah kemunculan (frekuensi) dari setiap kategori
d. Menentukan kategori menggunakan nilai maksimal

Rumus perhitungan untuk *Naïve Bayes* sebagai berikut [8]:

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

dimana:

X : data dengan kelas yang belum diketahui
H : hipotesis data X merupakan suatu kelas spesifik
$P(H \mid X)$: probabilitas hipotesis H berdasarkan kondisi X
$P(H)$: probabilitas hipotesis H
$P(X \mid H)$: probabilitas X berdasarkan kondisi pada hipotesis H
$P(X)$: probabilitas hipotesis X

2.4. Algoritma K-NN (*K-Nearest Neighbor*)

Algoritma *K-Nearest Neighbor* (K-NN) digunakan untuk melakukan klasifikasi pada objek yang didasari oleh data pembelajaran dengan jarak paling dekat terhadap objek tersebut [9]. Fitur-fitur yang relevan sangat memengaruhi ketepatan algoritma ini [10]. Tahapan dalam menggunakan algoritma K-NN, sebagai berikut:

a. Menentukan terlebih dahulu banyaknya tetangga k (sebaiknya ganjil)
b. Menghitung jarak pada data untuk dilakukannya perbandingan dengan dataset *training* menggunakan persamaan jarak Euclidean

$$dist\ (p,q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \quad (2)$$

dimana:

$dist(p,q)$: jarak antara p dan q
pi : nilai ke-i dari data p
qi : nilai ke-i dari data q

c. Mengatur urutan berdasarkan kecil ke besar kemudian pilih himpunan k yang paling sedikit pada dataset terkecil

ISSN : 2685-1814 (*Print*)
ISSN : 2685-7677 (*Online*)

d. Menentukan bahwa jawaban dengan data yang akan diprediksi adalah kelompok data yang memiliki jumlah k pertama dari kumpulan data terbesar
e. Gunakan titik pertimbangan untuk menetapkan kelas-kelas terdekat

2.5. Evaluasi Hasil

Confusion matrix merupakan tabel yang menyatakan jumlah suatu data uji yang diklasifikasikan secara salah maupun secara benar. *Confusion matrix* adalah sebuah matriks yang menampilkan visualisasi kinerja dari algoritma klasifikasi menggunakan data dalam matriks. Data dalam matriks membagi klasifikasi prediksi dalam empat bentuk yaitu *True Positif* (TP), *True Negatif* (TN), *False Positif* (FP), dan *False Negatif* (FN). Bentuk *confusion matrix* untuk klasifikasi dengan dua kelas, terdapat pada Tabel 2 berikut ini [11].

Tabel 2. *Confusion matrix*

Kelas	Prediksi YES	Prediksi NO	Total
Aktual YES	*True Positif* (TP)	*False Negatif* (FN)	*Positif* (P)
Aktual NO	*False Positif* (FP)	*True Negatif* (TN)	*Negatif* (N)
Total	P’	N’	P+N

Rumus mencari Presisi:

$$presisi = \frac{TP}{TP + FP} \tag{3}$$

Rumus mencari Recall:

$$recall = \frac{TP}{TP + FN} \tag{4}$$

Rumus mencari Akurasi:

$$akurasi = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

3. HASIL DAN PEMBAHASAN

3.1. Hasil

Klasifikasi dataset pada penyakit hati dengan menerapkan 2 algoritma yaitu *Naïve Bayes* dan algoritma K-NN memberikan hasil yang tidak sama. Berikut hasil *Confusion Matrix* nya.

Tabel 3. *Naïve Bayes*

Kelas		Nilai Aktual	
		Ada Penyakit	Tidak Ada Penyakit
Nilai Prediksi	Ada Penyakit	390	109
	Tidak Ada Penyakit	74	452

Berdasarkan Tabel 3 di atas dapat dilihat bahwa algoritma *Naïve Bayes* menyatakan bahwa sebanyak 390 orang terkena penyakit diprediksi sebagai terkena penyakit, sebanyak 109 orang tidak terkena penyakit diprediksi sebagai terkena penyakit, sebanyak 74 orang terkena penyakit diprediksi sebagai tidak terkena penyakit, dan sebanyak 452 orang tidak terkena penyakit diprediksi sebagai tidak terkena penyakit.

Tabel 4. *K-Nearest Neigbhor*

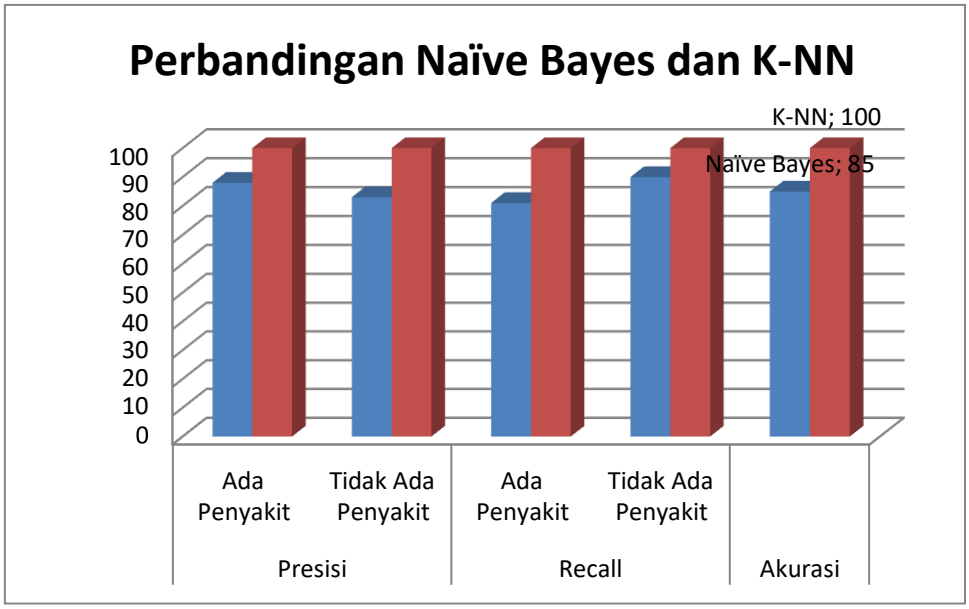
Kelas		Nilai Aktual	
		Ada Penyakit	Tidak Ada Penyakit
Nilai Prediksi	Ada Penyakit	429	70
	Tidak Ada Penyakit	95	431

Sedangkan berdasarkan algoritma *K-Nearest Neigbhor* pada Tabel 4 menyatakan bahwa sebanyak 429 orang terkena penyakit diprediksi sebagai terkena penyakit, sebanyak 70 orang tidak terkena penyakit diprediksi sebagai terkena penyakit, sebanyak 95 orang terkena penyakit diprediksi sebagai

ISSN : 2685-1814 (*Print*)
ISSN : 2685-7677 (*Online*)

tidak terkena penyakit, dan sebanyak 431 orang tidak terkena penyakit diprediksi sebagai tidak terkena penyakit.

Pada penerapan algoritma *Naïve Bayes* diperoleh nilai akurasi mencapai sebesar 85% sedangkan algoritma *K-Nearest Neigbhor* memperoleh nilai akurasi sebesar 100%. Untuk nilai presisi terkena penyakit pada algoritma *Naïve Bayes* mencapai sebesar 88% dan untuk tidak terkena penyakit sebesar 83% sedangkan algoritma *K-Nearest Neigbhor* untuk terkena penyakit mencapai sebesar 100% dan untuk tidak terkena penyakit sebesar 100%. Nilai *recall* untuk terkena penyakit pada algoritma *Naïve Bayes* mencapai sebesar 81% dan untuk tidak terkena penyakit sebesar 90% sedangkan algoritma *K-Nearest Neigbhor* untuk terkena penyakit sebesar 100% dan untuk tidak terkena penyakit sebesar 100%.



Gambar 1. Nilai-nilai presisi, *recall* dan akurasi *Naives Bayes* dan K-NN

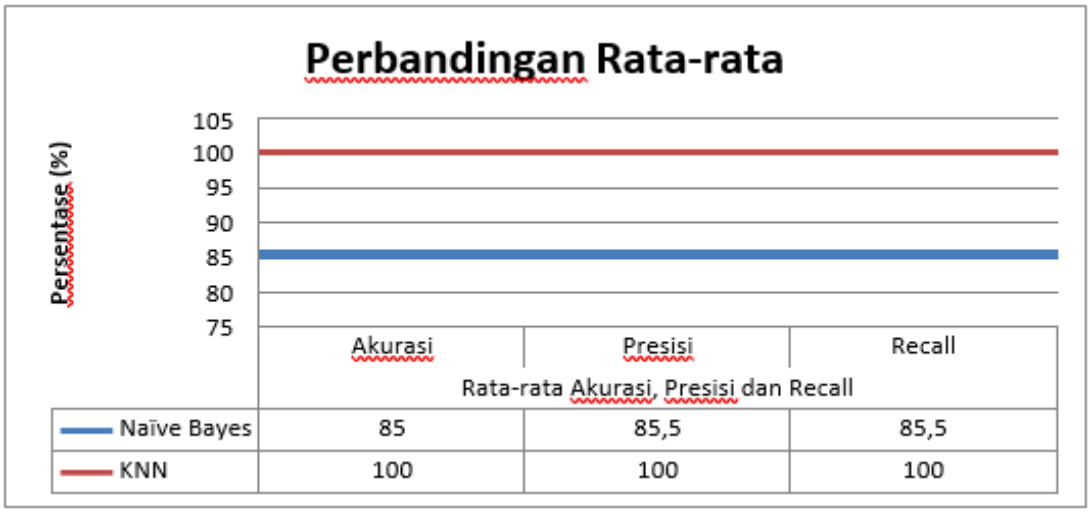
3.2. Perbandingan Hasil Kerja Kedua Metode

Hasil kerja atau prediksi yang didapatkan pada penggunaan 2 algoritma yaitu *Naïve Bayes* dan algortima *K- Nearest Neighbor* membuktikan bahwa dua algoritma tersebut baik dalam melakukan klasifikasi penyakit hati. Berikut ini ditampilkan perbandingan hasil rata-rata nilai presisi, *recall* dan akurasi.

Tabel 5. Perbandingan Rata-rata nilai presisi, *recall* dan akurasi kedua algoritma

Algoritma	Presisi	Recall	Akurasi
Naïve Bayes	85,5%	85,5%	85%
K-Nearest Neighbor	100%	100%	100%

ISSN : 2685-1814 (*Print*)
ISSN : 2685-7677 (*Online*)



Gambar 2. Perbandingan rata-rata presisi, *recall* dan akurasi

Berdasarkan Gambar 2 terlihat bahwa nilai Presisi dan Recall yang diperoleh algoritma *Naïve Bayes* dan algoritma *K-Nearest Neighbor* terdapat perbedaan yang cukup stabil sebagai prediksi penyakit hati karena algortima *Naïve Bayes* 85% keatas sedangkan algoritma *K-Nearest Neigbhor* mencapai 100% yang artinya memiliki perbedaan atau selisih sekitar sebesar 15%.

4. KESIMPULAN

Pada penelitian ini dapat diketahui bahwa pengklasifikasian dataset penyakit hati dengan menggunakan algoritma *Naïve Bayes* dan K-NN cukup baik. Terlihat dari nilai akurasi, presisi dan *recall* dari kedua algoritma yang melebihi nilai 80%. Algoritma *Naïve Bayes* memberikan rata-rata akurasi sebesar 85% yang bernilai lebih kecil disbandingkan dengan K-NN yang dapatmemberikan rata-rata bernilai sempurna yaitu 100%. Rata-rata nilai presisi dan recall dari Naïve *Bayes* bernilai sama yaitu 85,5%, sedangkan rata-rata nilai presisi dan *recall* dari K-NN bernilai sempurna atau 100%. Mulai dari akurasi, presisi dan *recall* terbaik ada pada algoritma K-NN. Dapat disimpulkan bahwa penggunaan algoritma K-NN pada klasifikasi penyakit hati akan sangat lebih baik dibandingkan dengan menggunakan algoritma *Naïve Bayes*.

REFERENSI

[1] A. Rosida, "Pemeriksaan Laboratorium Penyakit Hati," *Berk. Kedokt.*, vol. 12, no. 1, pp. 123–131, 2016.
[2] A. Pujiyanta dan A. Pujiantoro, "Sistem Pakar Penentuan Jenis Penyakit Hati dengan Metode Inferensi Fuzzy Tsukamoto," *J. Inform.*, vol. 6, no. 1, pp. 617–629, 2012.
[3] F. Safithri, "Mekanisme Regenerasi Hati secara Endogen pada Fibrosis Hati," *Magna Med. Berk. Ilm. Kedokt. dan Kesehat.*, vol. 2, no. 4, pp. 9–26, 2018.
[4] A. I. Falatehan, N. Hidayat, dan K. C. Brata, "Sistem Pakar Diagnosis Penyakit Hati Menggunakan Metode Fuzzy Tsukamoto Berbasis Android," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 8, pp. 2373–2381, 2018.
[5] C. N. Prabiantissa, "Klasifikasi pada Dataset Penyakit HatiMenggunakan Algoritma Support Vector Machine, K-NN, dan Naive Bayes," in *Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika*, vol. 1, no. 1, pp. 263–268, 2021.
[6] R. G. et al Rafsanjani, "Diagnosis Penyakit Hati Menggunakan Metode Naive Bayes Dan Certainty Factor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 4478–4482, 2018.