

Model Data Mining untuk Perancangan Aplikasi Diagnostik Inflammatory Liver Disease

Rahma Aulia Siahaan^{*1}, Marnis Nasution², Mila Nirmala Sari Hasibuan³

^{1,2,3}Universitas Labuhan Batu, Rantau Prapat, Indonesia

Email: rahmaaulia773000@gmail.com^{*1}, marnisnst@gmail.com², milanirmalasari7@gmail.com³

Abstrak

Hati merupakan organ vital bagi manusia. Penyakit hati adalah gangguan pada setiap fungsi hati. Diagnosis dini penyakit hati sangat penting agar dapat diobati dan diobati dengan cepat. Di bidang medis, mendiagnosis penyakit radang hati menjadi hal yang agak sulit dilakukan. Namun, ada catatan medis yang menyimpan gejala pasien. Hal ini tentunya sangat menguntungkan bagi tenaga medis atau dokter. Mereka dapat menggunakan catatan medis sebelumnya sebagai bahan untuk membuat keputusan tentang diagnosis penyakit pasien. Teknik analisis manual konvensional yang selama ini digunakan sudah tidak efektif lagi untuk diagnosis. Seiring dengan perkembangan sistem berbasis pengetahuan medis, tuntutan penggunaan sistem pengetahuan berbasis komputer sebagai teknik analisis dalam mendiagnosis penyakit menjadi semakin penting. Dalam studi ini, peneliti akan menerapkan dan membandingkan beberapa metode klasifikasi data mining, antara lain algoritma C4.5, Naïve Bayes, dan k-Nearest Neighbor untuk mendiagnosis penyakit radang hati, kemudian membandingkan mana dari ketiga metode tersebut yang paling akurat. Berdasarkan hasil pengukuran performansi ketiga model menggunakan metode Cross Validation, Confusion Matrix dan ROC Curve, diketahui bahwa metode C4.5 merupakan metode terbaik dengan akurasi 70,99% dan under the curve (AUC) nilai 0,950, kemudian metode k-Nearest Neighbor dengan akurasi 67,19% dan nilai under the curve (AUC) 0,873, kemudian metode nave Bayes dengan tingkat akurasi 66,14% dan nilai under the curve (AUC) sebesar 0,742.

Kata kunci: C4.5, nave bayes, k-Nearest Neighbor, matriks konfusi, Kurva ROC

Abstract

The liver is a vital organ for humans. Liver disease is a disorder of any liver function. Early diagnosis of liver disease is very important so that it can be treated and treated quickly. In the medical field, diagnosing inflammatory liver disease is a rather difficult thing to do. However, there are medical records that store the patient's symptoms. This is certainly very beneficial for medical personnel or doctors. They can use previous medical records as material for making decisions about the patient's disease diagnosis. Conventional manual analysis techniques that have been used so far are no longer effective for diagnosis. Along with the development of medical knowledge-based systems, demands for the use of computer-based knowledge systems as analytical techniques in diagnosing diseases are becoming increasingly important. In this study, researchers will apply and compare several data mining classification methods, including the C4.5, Naïve Bayes, and k-Nearest Neighbor algorithms to diagnose inflammatory liver disease, then compare which of the three methods is the most accurate. Based on the results of measuring the performance of the three models using the Cross Validation, Confusion Matrix and ROC Curve methods, it is known that the C4.5 method is the best method with an accuracy of 70.99% and under the curve (AUC) a value of 0.950, then the k-Nearest Neighbor method with an accuracy of 67.19% and an under the curve (AUC) value of 0.873, then the Nave Bayes method with an accuracy rate of 66.14% and an under the curve (AUC) value of 0.742.

Keywords: C4.5, nave bayes, k-Nearest Neighbor, confusion matrix, ROC . Curve

1. PENDAHULUAN

Hati merupakan organ vital bagi manusia. Organ ini terletak di rongga perut kanan, tepatnya di bawah diafragma. Ada beberapa fungsi hati, antara lain sebagai penangkal dan penetral racun, mengatur sirkulasi hormon, mengatur komposisi darah yang mengandung lemak, gula, protein, dan zat lainnya.

Hati juga membuat empedu, zat yang membantu mencerna lemak. Penyakit hati merupakan gangguan pada setiap fungsi hati. Hati bertanggung jawab atas fungsi-fungsi penting dalam tubuh, di mana hilangnya fungsi-fungsi ini dapat menyebabkan kerusakan yang signifikan pada tubuh. Hati adalah satu-satunya organ dalam tubuh yang dapat dengan mudah menggantikan sel-sel yang rusak, tetapi jika sel-sel ini hilang, hati mungkin tidak dapat memenuhi kebutuhan tubuh.

Belakangan ini, di bidang medis, mendiagnosis penyakit radang hati menjadi hal yang agak sulit dilakukan. Namun, ada catatan medis yang menyimpan gejala pasien. Hal ini tentunya sangat menguntungkan bagi tenaga medis atau dokter. Mereka dapat menggunakan catatan medis sebelumnya sebagai bahan untuk membuat keputusan tentang diagnosis penyakit pasien.

Teknik analisis manual konvensional yang selama ini digunakan sudah tidak efektif lagi untuk diagnosis. Seiring dengan perkembangan sistem berbasis pengetahuan medis, tuntutan penggunaan sistem pengetahuan berbasis komputer sebagai teknik analisis dalam mendiagnosis penyakit menjadi semakin penting. Oleh karena itu, saat ini adalah waktu yang tepat untuk mengembangkan sistem pengetahuan berbasis komputer yang modern, efektif dan efisien dalam mendiagnosis penyakit (Neshat & Yaghoobi, 2009).

Penelitian ini sesuai dengan penelitian yang dilakukan oleh (Nahar & Ara, 2018) yang meneliti tentang Prediksi Penyakit Hati Dengan Menggunakan Teknik Pohon Keputusan Berbeda. Penelitian ini menggunakan beberapa algoritma pohon keputusan seperti J48, LMT, Random Forest, Random tree, REPTree, Decision Stump dan Hoeffding Tree untuk memprediksi penyakit liver pada stadium awal. Dari hasil analisis, Decision Stump memiliki performa yang lebih baik dibandingkan algoritma lainnya dan akurasi yang dicapai adalah 70,67%.

Penelitian ini terkait dengan penelitian sebelumnya yang dilakukan oleh (Thirunavukkarasu et al., 2018) yang meneliti tentang Prediksi Penyakit Hati Menggunakan Algoritma Klasifikasi. Dari percobaan diketahui bahwa model k-nearest neighbor, akurasi yang telah dihitung dan yang keluar adalah 73,97%. Model Regresi Logistik, akurasi telah dihitung dan keluar menjadi 73,97%. Mendukung model Vector Machine, akurasi telah dihitung dan menjadi 71,97%.

Penelitian yang dilakukan (Setiawati et al., 2019) yang meneliti diagnosis penyakit hati dengan metode Decision Tree. Dari uji akurasi diketahui bahwa akurasi metode Naïve Bayes adalah 72,67%.

Penelitian yang dilakukan (Prayoga, 2018) yang meneliti diagnosis penyakit hati dengan metode Naïve Bayes. Dari uji akurasi diketahui bahwa akurasi metode Naïve Bayes adalah 87,50%.

Dalam penelitian ini, peneliti akan menerapkan dan membandingkan beberapa metode klasifikasi data mining, antara lain Algoritma C4.5, Naïve Bayes, dan k-nearest neighbor untuk mendiagnosis penyakit inflamasi hati. Dari ketiga metode tersebut manakah yang paling akurat dalam mendiagnosis penyakit radang hati. Penelitian ini diharapkan dapat membantu petugas kesehatan dalam mendiagnosis penyakit liver trafficking secara dini, dan dengan menggunakan aplikasi ini diharapkan masyarakat umum dapat memprediksi apakah suatu penyakit radang hati merupakan penyakit hati. terdiagnosis atau tidak. Jika didiagnosis, orang sesegera mungkin berkonsultasi dan menghubungi dokter.

2. METODE PENELITIAN

2.1 Penyakit hati

Penyakit hati merupakan penyakit yang telah menjadi masalah nasional di semua negara, baik di negara berkembang seperti Indonesia maupun di negara maju. Penyakit ini dapat terjadi pada semua kelompok umur, mulai dari anak-anak, remaja, dewasa hingga orang tua. Statistik menunjukkan bahwa penyakit hati adalah salah satu penyebab utama kematian, baik di Amerika Serikat maupun di seluruh dunia. Orang dengan penyakit hati sulit dideteksi, terutama pada tahap awal penyakit. Hal ini karena pasien tidak merasakan gejala penyakit apapun dan seolah-olah hati berfungsi normal, meskipun sebagian dari hati telah rusak (Hannan et al., 2010).

2.2 Penambangan Data

Menurut Larose (2005) dalam (Pahlevi et al., 2018) dalam bukunya yang berjudul “Discovering Knowledge in Data: An Introduction to Data Mining”, data mining dibagi menjadi beberapa kelompok berdasarkan tugas/pekerjaan yang dapat dilakukan, seperti sebagai:

1. Deskripsi

Terkadang peneliti dan analis hanya ingin mencoba mencari cara untuk menggambarkan pola dan tren yang terkandung dalam data. Deskripsi pola tren sering memberikan penjelasan yang mungkin untuk suatu pola atau tren.

2. Perkiraan

Estimasi hampir sama dengan klasifikasi, hanya saja variabel target yang diestimasi lebih numerik

daripada kategoris. Model dibangun dengan menggunakan sederet data (rekaman) lengkap yang memberikan nilai variabel target sebagai nilai prediktif. Selanjutnya pada tinjauan selanjutnya, estimasi nilai variabel target dibuat berdasarkan nilai variabel prediksi.

3. Ramalan

Prediksi hampir sama dengan klasifikasi dan estimasi, hanya saja dalam prediksi nilai hasil yang akan datang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi juga dapat digunakan (dalam keadaan yang sesuai) untuk prediksi.

4. Klasifikasi

Dalam klasifikasi, ada variabel kategoris target. Misalnya, klasifikasi pendapatan dapat menjadi dibedakan menjadi tiga kategori, yaitu berpenghasilan tinggi, berpenghasilan menengah, dan berpenghasilan rendah.

5. Kekelompokan

Clustering adalah pengelompokan record, observasi, atau perhatian dan membentuk kelas objek yang memiliki kemiripan. Cluster adalah kumpulan record yang mirip satu sama lain dan memiliki record yang berbeda di cluster lain. Berbeda dengan klasifikasi, dalam clustering tidak ada variabel target. Clustering tidak mengklasifikasi, memperkirakan, atau memprediksi nilai variabel target, namun algoritma clustering mencoba membagi seluruh data ke dalam kelompok-kelompok yang serupa (homogen), dimana kesamaan record dalam satu kelompok akan menjadi nilai maksimum, sedangkan kesamaan dengan record dalam kelompok lain akan menjadi minimal nilainya.

6. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Salah satu implementasi dari asosiasi adalah analisis keranjang pasar atau algoritma apriori.

2.3 Algoritma C4.5

Menurut Han dan Kamber dalam (Amrin, 2018a) algoritma C4.5 merupakan struktur pohon yang di dalamnya terdapat node yang menggambarkan atribut, setiap cabang menggambarkan hasil atribut yang diuji, dan setiap daun menggambarkan kelas. Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, hingga tidak dapat dibagi lagi. Algoritma C4.5 menggunakan konsep information gain atau entropy reduction untuk memilih distribusi yang optimal. Adapun menurut (Prasetyo, 2014) dalam (Handrianto & Farhan, 2019) dijelaskan bahwa, "Algoritma C4.5 diperkenalkan oleh J. Ross Quinlan (1996) sebagai versi perbaikan dari ID3 (Iterative Dichotomiser 3).

peningkatan yang membedakan algoritma C4.5 dari ID3 adalah dapat menangani fitur tipe numerik, memangkas pohon keputusan, dan menurunkan kumpulan aturan. Algoritma C4.5 digunakan untuk pohon keputusan.

2.4 Naïve Bayes

Menurut (Kusrini & Luthfi, 2009) Klasifikasi Bayes adalah klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas suatu keanggotaan kelas.

Menurut Han dan Kamber dalam (Amrin, 2018) klasifikasi Bayes, juga dikenal sebagai Naïve Bayes, memiliki kemampuan yang sebanding dengan pohon keputusan dan jaringan saraf.

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \quad (1)$$

Ucapan:

Y = data dengan kelas yang tidak diketahui

x = hipotesis data y adalah kelas tertentu

P(x | y) = probabilitas hipotesis x berdasarkan kondisi y (probabilitas posteriori)

$P(x)$ = probabilitas hipotesis x (probabilitas sebelumnya)
 $P(y|x)$ = probabilitas y berdasarkan kondisi dalam hipotesis x $P(y)$ = probabilitas y
 Naive Bayes adalah penyederhanaan dari metode Bayes. Teorema Bayes disederhanakan menjadi:

$$P(x|y) = P(y|x) P(x) \quad (2)$$

2.5 Algoritma K-Nearest Neighbor

Algoritma K-Nearest Neighbor (K-NN) merupakan metode yang menggunakan algoritma Supervised (Han & Kamber, 2006). K-Nearest Neighbor (K-NN) adalah sekelompok pembelajaran berbasis instance. Algoritma ini juga merupakan teknik pembelajaran malas. KNN dilakukan dengan mencari kelompok k objek pada data latih yang paling dekat (mirip) dengan objek pada data baru atau data pengujian (Wu & Kumar, 2009).

2.6 Matriks Kebingungan

Matriks Kebingungan adalah alat visualisasi yang biasa digunakan dalam pembelajaran terawasi. Setiap kolom dalam matriks merupakan contoh dari kelas prediksi, sedangkan setiap baris mewakili suatu kejadian di kelas yang sebenarnya (Gorunescu, 2011).

E. Kurva ROC (Karakteristik Operasi Penerima)

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan matriks kebingungan. ROC adalah grafik dua dimensi dengan positif palsu sebagai garis horizontal dan positif benar sebagai garis vertikal (Vercellis, 2009).

$$\theta^r = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m \psi(x_i^r, x_j^r) \quad (3)$$

Dimana :

$$\psi(X,Y) = \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 0 & Y > X \end{cases} \quad (4)$$

Menurut (Gorunescu, 2011) akurasi kinerja AUC dapat diklasifikasikan menjadi lima kelompok, yaitu sebagai berikut:

- 0,90 – 1,00 = Klasifikasi Luar Biasa
- 0,80 – 0,90 = Klasifikasi Baik
- 0,70 – 0,80 = Klasifikasi Wajar
- 0,60 – 0,70 = Klasifikasi Buruk
- 0,50 – 0,60 = Kegagalan

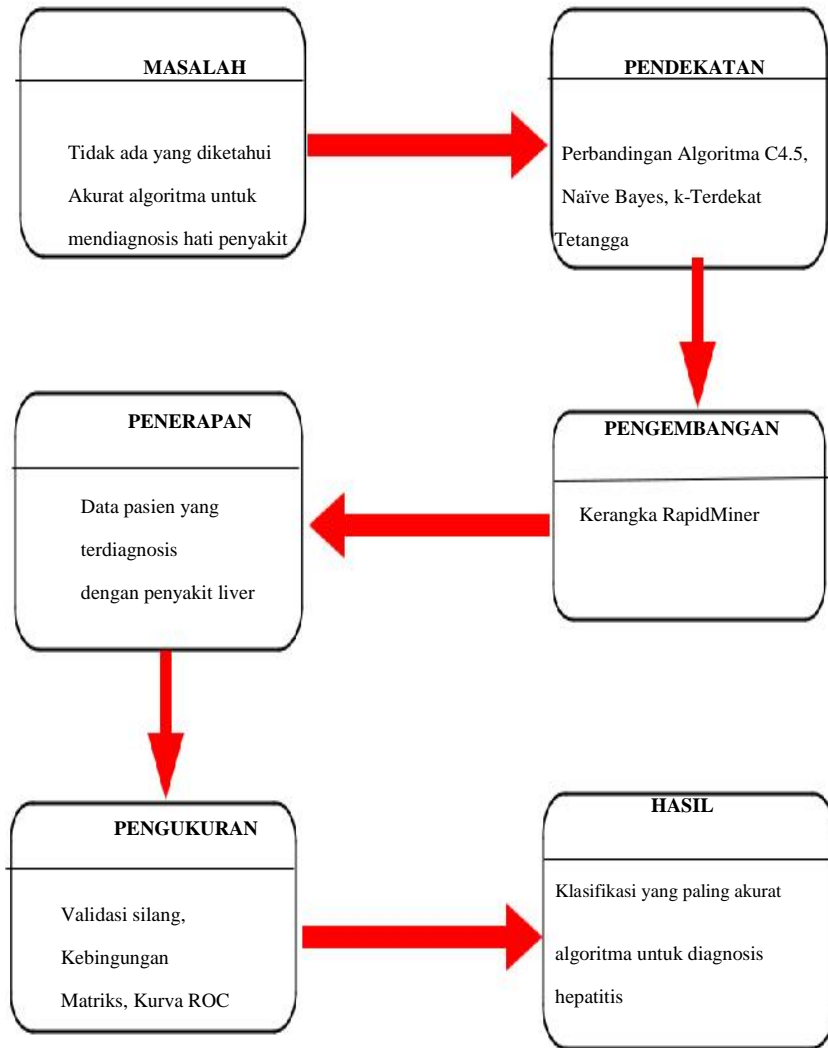
F. Rapid Miner

Menurut (Crc & Hofmann, 2014) dalam (Sumpena et al., 2019) Rapid Miner adalah sistem yang mendukung desain dan dokumentasi dari proses data mining secara keseluruhan. Ini bukan hanya seperangkat operator yang hampir komprehensif, tetapi juga struktur yang mengekspresikan kontrol proses.

3. METODE PENELITIAN

Penelitian ini terdiri dari beberapa tahapan seperti terlihat pada kerangka Gambar 1. Permasalahan dalam penelitian ini adalah belum diketahuinya algoritma yang akurat untuk mendiagnosis hepatitis. Untuk itu dibuat suatu pendekatan (model) yaitu algoritma C4.5, Naive Bayes, dan k-Nearest Neighbor untuk menyelesaikan masalah tersebut, kemudian dilakukan pengujian terhadap kinerja ketiga metode tersebut. Pengujian menggunakan metode Cross Validation, Confusion Matrix dan kurva ROC.

Untuk mengembangkan aplikasi (pengembangan) berdasarkan model yang dibuat digunakan Rapid Miner.



Gambar 1. Kerangka Pemecahan Masalah

4. HASIL DAN PEMBAHASAN

4.1 Analisis data

Dalam penelitian ini dataset yang digunakan adalah dari website UCI Machine Learning Repository yaitu Indian Liver Patient Dataset (ILPD). Dataset ini berisi data yang dikumpulkan dari pasien yang ada di timur laut Andhra Pradesh, India. Setelah dilakukan teknik preprocessing data, dataset tersebut berisi 414 pasien liver, sedangkan 165 pasien bukan penderita liver. Dataset ini memiliki 10 atribut dimana 9 atribut adalah atribut input sedangkan 1 atribut adalah output atau kelas, dan memiliki 579 record. Deskripsi atribut seperti yang ditunjukkan pada tabel 1 di bawah ini:

Tabel 1. Deskripsi Atribut

Atribut	Keterangan
Usia	Usia Pasien
TB	Bilirubin Total pasien
DB	Pasien Langsung
Alkphos	Bilirubin
Sersan	Fosfatase alkali
Astaga	Almin
Tp	Total Protein
ALB	Albumin
A/G/ Ratio	Albumin dan Globulin Jumlah Protein Albumin
Keals	Perbandingan Kelas Apakah pasien positif lever atau tidak

4.2 Pengujian Model

Penelitian ini dilakukan dengan menguji coba model yang diusulkan. Kemudian mengevaluasi dan memvalidasi model untuk menghasilkan nilai akurasi dan AUC. Pengujian menggunakan Rapidminer dengan operator cross-validation 10 kali lipat untuk mendapatkan akurasi dan hasil AUC untuk setiap algoritma yang diuji. Evaluasi dilakukan dengan Confusion Matrix dan ROC Curve atau Area Under Curve (AUC).

1. Matriks Kebingungan

Tabel 2 adalah matriks konfusi untuk algoritma C4.5. Diketahui dari 579 data, 407 data yang tergolong “Ya” diprediksi sesuai data sebenarnya, kemudian 7 data diprediksi “Tidak” tetapi ternyata “Ya”. Kemudian 4 data yang diklasifikasikan sebagai "Tidak" diprediksi sesuai, dan 161 data yang diprediksi "Ya" ternyata "Tidak".

Tabel 2. Model Matriks Kebingungan untuk Algoritma C4.5

Akurasi : 70,99% +/-1,98%(mikro:70,98%)			
	Benar	Benar	Kelas
	Iya	Tidak	presisi
Sebelum. Iya	407	161	71,65%
Sebelum. Tidak	7	4	36,36%
Ingatan Kelas	98,31%	2,42%	

Tabel 3. adalah matriks konfusi untuk algoritma Naïve Bayes. Diketahui dari 579 data, 270 data diklasifikasikan “Ya” persis sesuai dengan data sebenarnya, kemudian 144 data diprediksi “Tidak” tetapi ternyata “Ya”. Kemudian 113 data yang diklasifikasikan "Tidak" diprediksi sesuai, dan 52 data yang diprediksi "Ya" ternyata "Tidak".

Tabel 3. Model Matriks Kebingungan untuk Algoritma Naïve Bayes

Akurasi : 66,14% +/-4.55%(mikro:66,15%)			
	Benar	Benar	Kelas
	Iya	Tidak	presisi
Sebelum. Iya	270	52	83,85%
Sebelum. Tidak	144	113	43,97%
Tingkatan Kelas	65,22%	68,48%	

Tabel 4 adalah matriks konfusi untuk algoritma k-Nearest Neighbor (k-NN). Diketahui dari 579 data, 318 data diklasifikasikan "Ya" persis sesuai dengan data sebenarnya, kemudian 96 data diprediksi "Tidak" tetapi ternyata "Ya". Kemudian 71 data yang diklasifikasikan "Tidak" diprediksi sesuai, dan 94 data yang diprediksi "Ya" ternyata "Tidak".

Tabel 4. Model Matriks Kebingungan untuk Metode K-Nearest Neighbor (K-NN)

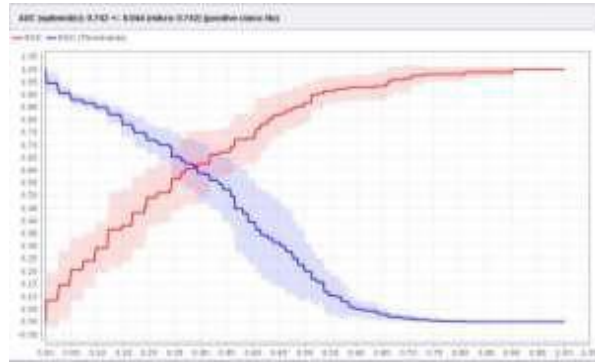
Akurasi : 67,19% +/- 3,79%(mikro:67,18%)			
	Benar	Benar	Kelas
	Iya	Tidak	presisi
Sebelum. Iya	318	94	77,18%
Sebelum. Tidak	96	71	42,51%
Ingatan Kelas	76,81%	43,03%	

Pada bagian ini peneliti memaparkan pembahasan penelitian yang hasil perhitungannya divisualisasikan dengan kurva ROC. Perbandingan ketiga metode perbandingan tersebut dapat dilihat pada Gambar 2 yang merupakan kurva ROC untuk algoritma C4.5.



Gambar 2. Kurva ROC dengan algoritma C4.5

Kurva ROC pada Gambar 2 mengungkapkan matriks kebingungan. Garis horizontal adalah positif palsu dan garis vertikal adalah positif benar. Selanjutnya kurva ROC untuk algoritma Naïve Bayes seperti terlihat pada Gambar 3 di bawah ini.



Gambar 3. Kurva ROC Algoritma Naïve Bayes



Gambar 4. Kurva ROC Algoritma k-Nearest Neighbor

Perbandingan hasil perhitungan nilai AUC untuk metode C4.5, naïve Bayes, dan k-Nearest Neighbor dapat dilihat pada Tabel 5.

Perbandingan akurasi dan nilai Kurva ROC atau AUC untuk algoritma C4.5, Naïve Bayes, dan k-Nearest Neighbor ditunjukkan pada tabel 5 di bawah ini. Metode C4.5 memiliki nilai akurasi tertinggi, disusul metode k-Nearest Neighbor, dan terendah adalah metode Naïve Bayes.

Tabel 5. Perbandingan nilai Akurasi dan AUC

Nilai	C4.5	Naif Bayes	k-NN
Ketepatan	70,99%	66,14%	67,19%
AUC	0,950	0,742	0,873

Tabel 5 membandingkan akurasi dan AUC masing-masing algoritma. Dapat dilihat bahwa nilai akurasi C4.5 untuk algoritma adalah yang tertinggi begitu juga dengan nilai AUCnya. Untuk klasifikasi data mining, nilai AUC dapat dibagi menjadi beberapa kelompok.

- a. 0,90-1,00 = klasifikasi sangat baik
- b. 0,80-0,90 = klasifikasi baik good
- c. 0,70-0,80 = klasifikasi sedang
- d. 0,60-0,70 = klasifikasi buruk
- e. 0,50-0,60 = klasifikasi salah

Berdasarkan pengelompokan di atas dan Tabel 5 dapat disimpulkan bahwa model algoritma C4.5 termasuk dalam kategori klasifikasi sangat baik, algoritma Naïve Bayes berada pada kategori klasifikasi cukup, dan k-Nearest Neighbor berada pada klasifikasi baik. kategori.

5. KESIMPULAN

Kesimpulan yang dapat diambil berdasarkan penelitian ini adalah kinerja model C4.5 untuk diagnosis penyakit hati memberikan tingkat akurasi sebesar 70,99% dengan nilai area under the curva (AUC) sebesar 0,950. Kinerja model nave Bayes memberikan tingkat akurasi kebenaran sebesar 66,14% dengan luas di bawah kurva (AUC) sebesar 0,742. Sedangkan performansi model k-Nearest Neighbo memberikan tingkat akurasi kebenaran sebesar 67,19% dengan area under the curve (AUC) sebesar 0,873 dan berdasarkan tingkat akurasi dan nilai area under the curve (AUC) , kinerja metode C4.5 adalah yang terbaik untuk mendiagnosis penyakit hati, diikuti dengan metode k-nearest neighbor, dan kemudian metode nave bayes.

DAFTAR PUSTAKA

- [1] Amrin, A. (2018a). Aplikasi Diagnosa Penyakit Tuberkulosis Menggunakan Algoritma Data Mining. *Jurnal Paradigma*, XX(2), 91–97.
- [2] Amrin, A. (2018b). Aplikasi Diagnosa Penyakit Tuberkulosis Menggunakan Algoritma Naive Bayes. *Jurikom*, 5(5), 498–502.
- [3] Gorunescu, F. (2011). *Data Mining: Konsep, Model, dan Teknik*. Peloncat.
- [4] Handrianto, Y., & Farhan, M. (2019). C.45 Algoritma Klasifikasi Penyebab Tanah Longsor. *Sinkron*, 4(1), 120–127. <https://doi.org/10.33395/sinkron.v4i1.10154>
- [5] Hannan, A., Manza, R., & Remteke, R. (2010). Jaringan Saraf Regresi Umum dan Fungsi Basis Radial untuk Diagnosis Penyakit Jantung. *Jurnal Internasional Aplikasi Komputer (0975-8887)*, 7 (13), 7-13.
- [6] Kusriani, & Luthfi, E. . (2009). *Algoritma Data Mining*. Penerbitan Andi.
- [7] Nahar, N., & Ara, F. (2018). Prediksi Penyakit Hati dengan Menggunakan Berbagai Teknik Pohon Keputusan. *Internasional Jurnal Proses Data Mining & Manajemen Pengetahuan*, 8(2), 01–09. <https://doi.org/10.5121/ijdkp.2018.8201>
- [8] Neshat, M., & Yaghoobi, M. (2009). Merancang Sistem Pakar Fuzzy Mendiagnosis Tingkat Intensitas Hepatitis B dan Membandingkannya dengan Sistem Fuzzy Neural Network Adaptif. *Prosiding Kongres Dunia tentang Teknik dan Ilmu Komputer 2009, Vol II, WCECS 2009*, ISBN:978-988-18210-2-7, 1–6.
- [9] Pahlevi, O., Sugandi, A., & Sintawati, ID (2018). Penerapan Algoritma Apriori Dalam Pengendalian Kualitas Produk. *Sinkron*, 3(1), 272–278.
- [10] Prayoga, ND (2018). Sistem Diagnosis Penyakit Hati Menggunakan Metode Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(8), 2666–2671.
- [11] Pusporani, E., Qomariyah, S., & Irhamah. (2019). Klasifikasi Pasien Penyakit Hati dengan Pendekatan Machine Learning. *Inferensi*, 2(1), 25–32. <https://doi.org/10.12962/j27213862.v2i1.6810>
- [12] Setiawati, I., Wibowo, AP, & Hermawan, A. (2019). Implementasi Decision Tree Untuk Mendiagnosis Penyakit Hati. *JOISME: Jurnal Manajemen Sistem Informasi*, 1(1), 13–17.
- [13] Sumpena, Akbar, Y., Nirat, & Henky, M. (2019). Perbandingan C4 . 5 Algoritma dan Nave Bayes untuk Informasi Terakhir pada Pasien ICU. *Sinkron*, 4(1), 88–94.
- [14] Thirunavukkarasu, K., Singh, AS, Irfan, M., & Chowdhury, A. (2018). Prediksi penyakit hati menggunakan Algoritma klasifikasi. *Konferensi Internasional ke-4 tentang Komunikasi dan Otomasi Komputasi 2018, ICCCA 2018*, 1(1), 1-3. <https://doi.org/10.1109/CCAA.2018.8777655>
- [15] Vercellis, C. (2009). *Business Intelligent: Data Mining dan Optimalisasi untuk Pengambilan Keputusan*. John Willey & Sons, Ltd.
- [16] Wu, X., & Kumar, V. (2009). *Sepuluh Algoritma Teratas dalam Data Mining*. CRC Pers.