Komparasi Penerapan Adaboost Pada K-NN Dan Decision Tree Untuk Prediksi Penyakit Hati

Hafsah Mukaromah*[1], Ratnasari Ratnasari[2]
[1]Program Studi S1 Farmasi, Universitas Aisyah Pringsewu; Jl. A Yani No. 1 A Tambak Rejo, Wonodadi, Kec. Pringsewu, Kabupaten Pringsewu, Lampung, 35372, Indonesia, telp (0729) 7081587
[2]Program Studi S1 Rekayasa Perangkat Lunak, Universitas Aisyah Pringsewu; Jl. A Yani No. 1 A Tambak Rejo, Wonodadi, Kec. Pringsewu, Kabupaten Pringsewu, Lampung, 35372, Indonesia, telp (0729) 7081587
e-mail: ***[1]hafsahmukaromah@aisyahuniversity.ac.id**, [2]ratnasari@aisyahuniversity.ac.id

Abstrak

Hati merupakan organ vital manusia yang berperan penting dalam proses detoksifikasi, pengaturan kadar kolesterol, serta berbagai aktivitas metabolisme tubuh. Gangguan pada fungsi hati dapat memicu berbagai penyakit seperti hepatitis, kanker hati, sirosis, dan jenis penyakit hati lainnya. Di Indonesia, sekitar 0,6% penduduk teridentifikasi mengidap hepatitis, meskipun program imunisasi HB 0–4 telah diupayakan oleh Kementerian Kesehatan. Penyakit liver menjadi isu kesehatan yang cukup umum, dengan data WHO mencatat angka kematian akibat penyakit ini mencapai 1,2 juta jiwa setiap tahunnya di kawasan Asia Tenggara dan Afrika. Kesadaran akan pentingnya deteksi dini gejala penyakit hati menunjukkan perlunya sistem prediksi yang mampu mengidentifikasi potensi penyakit ini secara akurat. Penelitian ini mengadopsi pendekatan machine learning menggunakan algoritma klasifikasi K-Nearest Neighbor (K-NN) dan Decision Tree, serta memperkuat performanya dengan teknik ensemble learning Adaboost. Berdasarkan hasil evaluasi, penggunaan Adaboost mampu meningkatkan akurasi algoritma K-NN hingga 95,77% dan meningkatkan akurasi Decision Tree hingga 100%. Walaupun peningkatan pada K-NN cukup signifikan, penerapan Adaboost tidak menunjukkan dampak yang besar terhadap Decision Tree. Penelitian ini mengindikasikan bahwa metode Adaboost efektif dalam meningkatkan performa klasifikasi penyakit hati, khususnya pada algoritma K-NN.

Kata kunci*— Hati; Prediksi; K-NN; Decision Tree; Adaboost.*

Abstract

The liver is a vital human organ that plays a crucial role in detoxification, cholesterol regulation, and various metabolic activities within the body. Impairment of liver function can lead to several diseases such as hepatitis, liver cancer, cirrhosis, and other liver-related conditions. In Indonesia, approximately 0.6% of the population is identified as having hepatitis, despite the implementation of the HB 0–4 immunization program by the Ministry of Health. Liver disease is a common public health issue, with WHO data reporting an annual death toll of 1.2 million people due to liver-related illnesses in Southeast Asia and Africa. The importance of early detection of liver disease symptoms highlights the need for a predictive system capable of accurately identifying individuals at risk. This study employs a machine learning approach using K-Nearest Neighbor (K-NN) and Decision Tree classification algorithms, enhanced by the application of the Adaboost ensemble learning technique to optimize their performance. Evaluation results show that Adaboost improves the accuracy of the K-NN algorithm to 95.77% and the accuracy of the Decision Tree to 100%. Although the improvement in K-NN is quite

significant, Adaboost does not have a substantial impact on the accuracy of the Decision Tree. This research indicates that the Adaboost method is effective in enhancing the classification performance for liver disease, particularly when applied to the K-NN algorithm.

Keywords— *Liver; Prediction; K-NN; Decision Tree; Adaboost*

1. PENDAHULUAN

Hati adalah organ yang sangat penting dalam tubuh manusia, yang memiliki peran utama dalam proses detoksifikasi untuk membersihkan racun dari dalam tubuh serta mengatur kadar kolestrol dan lemak. [1] Selain itu hati juga berfungsi sebagai pusat metabolisme yang mendukung kebutuhan energi otak dan menyaring darah dari zat berbahaya yang berasal dari usus. Gangguan pada fungsi hati dapat menyebabkan berbagai penyakit, seperti hepatitis, kanker hati, sirosis dan gangguan hati lainnya. [2]

Di Indonesia sekitar 0,6% di antaranya menderita penyakit hepatitis. Kementerian Kesehatan telah berusaha keras untuk mencegah penyakit ini melalui program imunisasi HB 0-4 yang dimulai sejak bayi. Berdasarkan data imunisasi dari Kementerian Kesehatan, capaian program imunisasi mengalami peningkatan yang signifikan dan telah melebihi target yang ditetapkan oleh pemerintah. Namun, peningkatan tersebut belum efektif dalam menurunkan jumlah penderita hepatitis di Indonesia. [3] Organ hati juga rentan terkena penyakit yang dapat mengganggu fungsi normalnya, bahkan dapat menyebabkan kematian. Penyakit liver merupakan kondisi yang sudah lama ada dan umum di masyarakat. Menurut data dari World Health Organization (WHO) menunjukkan bahwa setiap tahunnya hampir 1,2 juta orang, terutama di wilayah Asia Tenggara dan Afrika, meninggal akibat penyakit liver. Faktor penyebab termasuk kelainan hati bawaan, infeksi virus atau bakteri, kecanduan alkohol, merokok aktif, pola hidup yang buruk, dan faktor lainnya. [4] Gejala penyakit liver dapat bersifat klinis atau fisik. Gejala klinis dapat teridentifikasi melalui pemeriksaan gejala yang dirasakan pasien, sementara gejala fisik dapat diidentifikasi melalui pemeriksaan fisik. Namun, dalam beberapa kasus, pasien tidak menyadari bahwa mereka mengidap penyakit hati. Bahkan tenaga medis, termasuk dokter, dapat mengalami kesulitan dalam mendeteksi gejala penyakit hati pada tahap awal, meskipun penyakit tersebut telah menyebar. [5]

Pengenalan terhadap tanda-tanda awal penyakit hati sangat penting agar individu dapat memperoleh pengobatan yang tepat serta meningkatkan peluang keselamatan hidup. Untuk mengatasi permasalahan tersebut, diperlukan suatu sistem yang mampu memprediksi kemungkinan seseorang mengidap penyakit hati atau tidak. Salah satu pendekatan yang relevan dalam mengatasi permasalahan prediksi penyakit adalah *machine learning*, yaitu suatu metode kecerdasan buatan yang memungkinkan komputer memperoleh pengetahuan dan kemampuan membuat keputusan berdasarkan pola yang dipelajari dari data historis yang tersedia.[6] Dalam machine learning, terdapat beberapa algoritma klasifikasi yang digunakan untuk memprediksi penyakit. K-Nearest Neighbor (KNN) merupakan salah satu algoritma klasifikasi yang banyak diterapkan dalam berbagai studi karena kemampuannya dalam mengelompokkan data berdasarkan kedekatan jarak antar titik. Meskipun sederhana untuk diimplementasikan, algoritma ini dapat memberikan akurasi yang baik. Namun, kelemahan KNN terletak pada penentuan nilai k, dimana jika nilai k terlalu besar, hasil klasifikasinya menjadi tidak jelas, sedangkan jika nilai k terlalu kecil, seperti k=1, hasil klasifikasinya menjadi kaku karena kurangnya variasi pilihan. Oleh karena itu, penelitian untuk menentukan nilai k yang optimal diperlukan.[6] Algoritma *Decision Tree* merupakan salah satu teknik klasifikasi yang memanfaatkan struktur pohon, di mana setiap node merepresentasikan atribut, sedangkan setiap daun menggambarkan kelas atau kategori hasil. [10] Kelebihan dari decision tree adalah modelnya yang sederhana dan mudah dimengerti karena direpresentasikan dalam bentuk pohon dengan cabang-cabang, mempermudah interpretasi. Decision tree juga memiliki akurasi yang

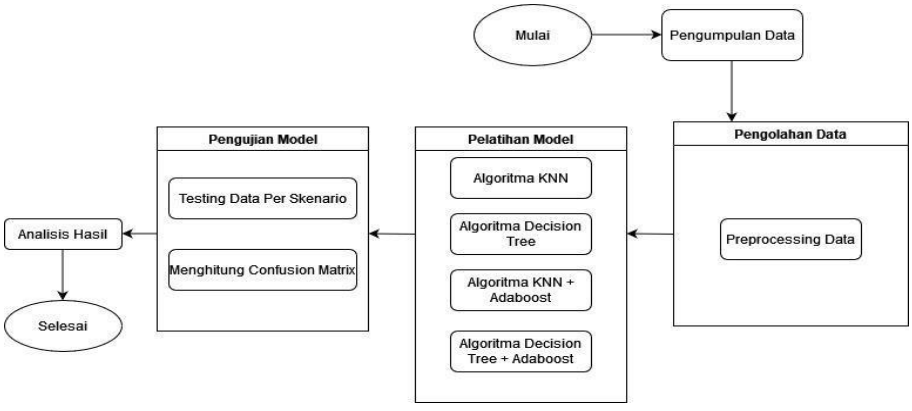
tinggi berdasarkan penelitian sebelumnya, mendukung penerapannya. [11] Sedangkan Adaboost, singkatan dari Adaptive Boosting, termasuk dalam Ensemble Methods/Boosting Methods yang umum digunakan. Secara umum, Adaboost membangun sejumlah weak learners yang tidak berkorelasi, lalu menggabungkan prediksinya. Dalam aplikasinya, Adaboost dapat dikombinasikan dengan algoritma lain untuk mengoptimalkan performa yang dihasilkan. [7] Penelitian sebelumnya membandingkan performa algoritma SVM, KNN, dan Naïve Bayes, dan hasilnya menunjukkan bahwa dari ketiga algoritma tersebut, SVM memiliki kinerja rata-rata paling baik dengan tingkat akurasi mencapai 82,36%. [2] Penelitian berikutnya, yang dilakukan oleh [1] dengan judul "Implementasi Decision Tree Untuk Mendiagnosis Penyakit Liver," menunjukkan bahwa algoritma Decision Tree mencapai tingkat akurasi sebesar 72,67% dalam mengklasifikasikan penyakit liver menggunakan Dataset ILPD. Penelitian ketiga, yang dilakukan oleh [5] dengan judul "Klasifikasi Pasien Penderita Penyakit Liver Dengan Pendekatan Machine Learning," membandingkan 5 algoritma klasifikasi, yaitu Regresi Logistik, Decision Tree, Naïve Bayes, SVM, dan K-NN. Berdasarkan akurasi dan presisi, algoritma SVM memberikan hasil terbaik, tetapi dari segi nilai recall, metode K-Nearest Neighbor memberikan hasil terbaik.

Dalam penelitian ini, digunakan algoritma klasifikasi *K-Nearest Neighbor* (KNN) dan *Decision Tree*, yang selanjutnya dioptimalkan dengan menerapkan teknik *ensemble learning* Adaboost. Penelitian ini bertujuan untuk mengatasi permasalahan ketidakakuratan dalam prediksi serta memverifikasi bahwa algoritma yang digunakan mampu menghasilkan prediksi dengan tingkat akurasi yang tinggi.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Dataset dalam penelitian ini diperoleh dari situs Kaggle dengan jumlah 615 data dan dianalisis menggunakan tools RapidMiner. Atribut yang digunakan sebanyak 12 atribut prediktor yaitu Age, Sex, ALB, ALP, ALT, AST, BIL, CHE, CHOL, CREA, GGT, PROT dan 1 atribut target/label yaitu Category. Bab ini akan membahas tahapan dalam proses penelitian yang akan dilakukan. Tujuan utamanya adalah untuk menganalisis dan mencari pola data yang akan digunakan sebagai dataset. Hal ini akan membantu mempermudah penelitian dan menjalankannya secara sistematis sesuai dengan tujuan yang diinginkan. Untuk mencapai tujuan tersebut, maka tahapan penelitian dirancang sebagai berikut:

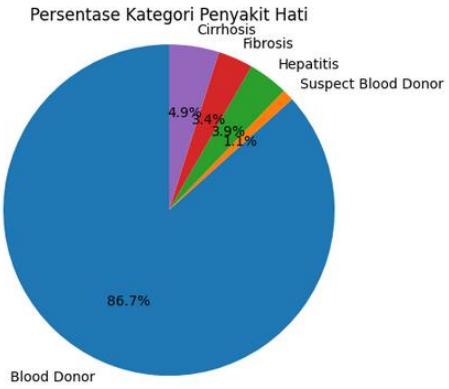


Gambar 1. Diagram Alur Penelitian

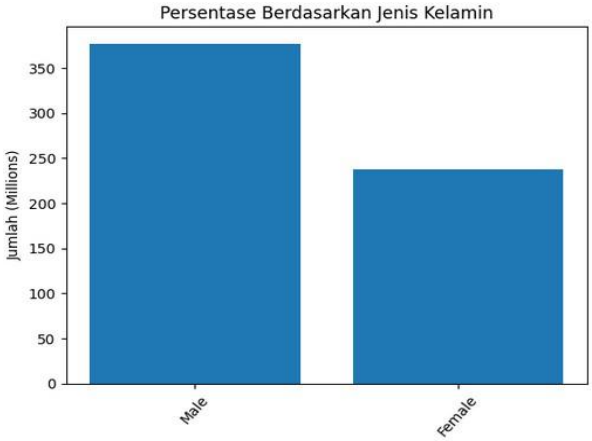
Berdasarkan diagram alur penelitian diatas, penelitian ini memiliki beberapa tahapan yang dilalui, yaitu:

1. Tahap Pengumpulan Data
   Pengumpulan data merupakan tahap awal dalam proses analisis data dan merupakan

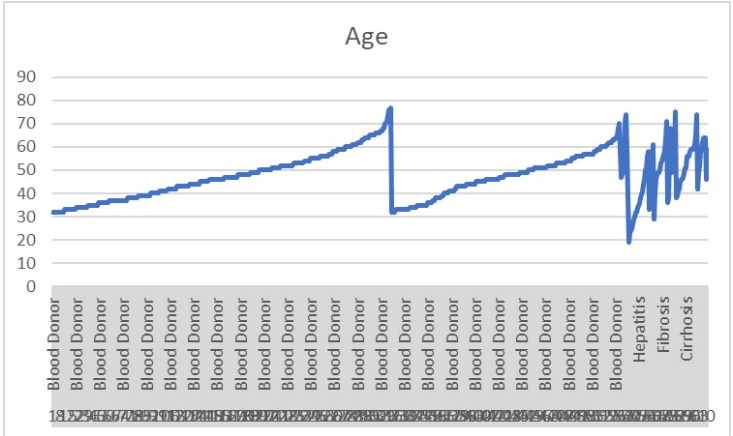
kunci untuk menjamin kesuksesan proyek. Dataset yang digunakan berasal dari situs Kaggle dan terdiri dari 615 data yang diolah menggunakan tools RapidMiner. Terdapat 12 atribut yang digunakan dalam analisis, yaitu Age, Sex, ALB, ALP, ALT, AST, BIL, CHE, CHOL, CREA, GGT, dan PROT.



Gambar 2. Persentase Kategori Penyakit Hati



Gambar 3. Persentase Berdasarkan Jenis Kelamin



Gambar 4. Persentase Berdasarkan Usia

2. Tahap Pengolahan Data/Preprocessing Data.

Pengolahan data atau preprocessing data adalah langkah kritis dalam analisis data yang bertujuan untuk membersihkan, mengorganisir, dan menyiapkan data mentah agar dapat digunakan secara efektif dalam proses analisis lebih lanjut. Pada dataset penyakit hati terdapat 10 data bernilai NA atau tidak konsisten yaitu pada atribut CHOL dan 21 data kosong/missing value pada atribut ALB, ALT, ALP, dan PROT. Data yang bernilai NA akan digantikan dengan nilai rata-rata dari atribut CHOL. Sedangkan 21 data missing value akan dilakukan replace missing value dengan mengganti nilai yang hilang menggunakan nilai rata-rata dari masing-masing atribut, yaitu ALB, ALT, ALP, dan PROT.

3. Tahap Pelatihan Model.

Pada tahap pelatihan model dilakukan pengujian algoritma klasifikasi dan teknik optimasi. Penelitian ini menggunakan algoritma klasifikasi K-Nearest Neighbor dan Decision Tree dengan penerapan teknik optimasi Adaboost. Oleh karena itu, dihasilkan empat proses klasifikasi, yaitu klasifikasi menggunakan algoritma K-Nearest Neighbor, Decision Tree, K-Nearest Neighbor dengan Adaboost, serta Decision Tree dengan Adaboost.

4. Tahap Pengujian Model

Pada fase ini, dilakukan pengujian terhadap algoritma klasifikasi yang digunakan. Sebelum pengujian, data dibagi menjadi dua bagian, yaitu data latih dan data uji. Evaluasi model dilakukan dengan menggunakan metode *Confusion Matrix*. Untuk menilai kinerja model berdasarkan *Confusion Matrix*, digunakan metrik evaluasi seperti precision, specificity, sensitivity, dan F-score. Sementara itu, untuk mengukur akurasi model, diterapkan metode *k-fold cross-validation* dengan menggunakan 10 sampel data yang diambil secara acak. Model dievaluasi dengan memilih model yang memiliki nilai optimal, yang diperoleh dari rata-rata akurasi hasil pengujian pada 10 sampel tersebut.

5. Tahap Analisis Hasil

Pada fase analisis hasil, dilakukan perbandingan terhadap setiap skenario proses yang telah dilakukan pada tahap pelatihan model sebelumnya, dengan tujuan untuk menentukan hasil pengujian yang paling optimal. Perbandingan ini melibatkan penilaian terhadap nilai presisi, spesifisitas, sensitivitas, f1-score, dan akurasi dari masing-masing skenario proses.

3. HASIL DAN PEMBAHASAN

3.1 Persiapan Data

Data yang digunakan dalam penelitian ini berasal dari dataset yang diunduh melalui situs Kaggle yaitu dataset penyakit hati, jumlah record data 615 data terdiri dari 12 atribut prediktor dan 1 atribut sebagai label. Kumpulan data ini digunakan untuk memprediksi apakah seorang pasien menderita penyakit hati atau tidak. Data penyakit hati terdapat pada gambar 5 dibawah ini:

CHOL	id	Category	Age	Sex	ALB	ALP	ALT	AST	BIL	CHE	CREA	GGT	PROT
3,2	1,0	0=Blood D	32,0	m	385,0	525,0	77,0	221,0	75,0	693,0	106,0	121,0	69,0
4,8	2,0	0=Blood D	32,0	m	385,0	703,0	18,0	247,0	39,0	1117,0	74,0	156,0	765,0
5,2	3,0	0=Blood D	32,0	m	469,0	747,0	362,0	526,0	61,0	884,0	86,0	332,0	793,0
4,7	4,0	0=Blood D	32,0	m	432,0	52,0	306,0	226,0	189,0	733,0	80,0	338,0	757,0
4,3	5,0	0=Blood D	32,0	m	392,0	741,0	326,0	248,0	96,0	915,0	76,0	299,0	687,0
6,1	6,0	0=Blood D	32,0	m	416,0	433,0	185,0	197,0	123,0	992,0	111,0	91,0	74,0
4,8	7,0	0=Blood D	32,0	m	463,0	413,0	175,0	178,0	85,0	701,0	70,0	169,0	745,0
4,6	8,0	0=Blood D	32,0	m	422,0	419,0	358,0	311,0	161,0	582,0	109,0	215,0	671,0
4,1	9,0	0=Blood D	32,0	m	509,0	655,0	232,0	212,0	69,0	869,0	83,0	137,0	713,0
4,5	10,0	0=Blood D	32,0	m	424,0	863,0	203,0	20,0	352,0	546,0	81,0	159,0	699,0
3,6	11,0	0=Blood D	32,0	m	443,0	523,0	217,0	224,0	172,0	415,0	78,0	241,0	754,0
4,3	12,0	0=Blood D	33,0	m	464,0	682,0	103,0	20,0	57,0	736,0	79,0	187,0	686,0
5,4	13,0	0=Blood D	33,0	m	363,0	786,0	236,0	22,0	7,0	856,0	78,0	194,0	687,0
3,4	14,0	0=Blood D	33,0	m	39,0	517,0	159,0	24,0	68,0	646,0	65,0	7,0	704,0
5,0	15,0	0=Blood D	33,0	m	387,0	398,0	225,0	23,0	41,0	463,0	63,0	152,0	719,0
4,4	16,0	0=Blood D	33,0	m	418,0	65,0	331,0	38,0	66,0	883,0	71,0	24,0	727,0
5,2	17,0	0=Blood D	33,0	m	409,0	73,0	172,0	229,0	10,0	698,0	90,0	147,0	724,0
5,5	18,0	0=Blood D	33,0	m	452,0	883,0	324,0	312,0	101,0	978,0	102,0	485,0	765,0
5,5	19,0	0=Blood D	33,0	m	366,0	571,0	389,0	403,0	249,0	962,0	112,0	276,0	693,0
4,1	20,0	0=Blood D	33,0	m	42,0	631,0	326,0	349,0	112,0	701,0	105,0	191,0	681,0

Gambar 5. Dataset Penyakit Hati

3.2 Pengujian Model

3.2.1 Algoritma K-NN

Pengujian model pertama dilakukan menggunakan algoritma K-Nearest Neighbor dengan metode *Confusion Matrix* untuk mengetahui nilai kelas sebenarnya dan nilai kelas prediksi dari dataset penyakit hati. Pasien yang dinyatakan memiliki riwayat pendonor darah (blood donor) sebanyak 532, pasien yang diduga berpotensi menjadi pendonor darah (suspect blood donor) sebanyak 1 orang, pasien yang menderita penyakit hepatitis sebanyak 13, pasien yang menderita Fibrosis ada 8, dan pasien yang menderita sirosis berjumlah 18 orang. Berikut tabel *Confusion Matrix* dari hasil pengujian menggunakan algoritma K-Nearest Neighbor.

	true 0=Blood D...	true 0s=suspec...	true 1=Hepatitis	true 2=Fibrosis	true 3=Cirrhosis	class precision
pred. 0=Blood ...	532	5	1	2	3	97.97%
pred. 0s=susp...	0	1	0	0	0	100.00%
pred. 1=Hepatitis	0	1	13	9	3	50.00%
pred. 2=Fibrosis	0	0	5	8	6	42.11%
pred. 3=Cirrhos...	1	0	5	2	18	69.23%
class recall	99.81%	14.29%	54.17%	38.10%	60.00%	

Gambar 6. Tabel Confusion Matrix Algoritma K-NN

Selain melakukan perhitungan menggunakan tabel confusion matrix, pengujian juga dilakukan menggunakan metode *k-fold cross-validation* untuk mengetahui nilai akurasi dari algoritma K-Nearest Neighbor. Nilai akurasi yang diperoleh dari perhitungan *10-fold cross-validation* adalah sebesar 93,01%. Berikut ini merupakan hasil pengujian algoritma K-NN.

PerformanceVector (Performance) | KNNClassification (k-NN)

Table View | Plot View

accuracy: 93.01% +/- 1.70% (micro average: 93.01%)

	true 0=Blood D...	true 0s=suspec...	true 1=Hepatitis	true 2=Fibrosis	true 3=Cirrhosis	class precision
pred. 0=Blood ...	532	5	1	2	3	97.97%
pred. 0s=susp...	0	1	0	0	0	100.00%
pred. 1=Hepatitis	0	1	13	9	3	50.00%
pred. 2=Fibrosis	0	0	5	8	6	42.11%
pred. 3=Cirrhos...	1	0	5	2	18	69.23%
class recall	99.81%	14.29%	54.17%	38.10%	60.00%	

Gambar 7. Pengujian Algoritma K-NN

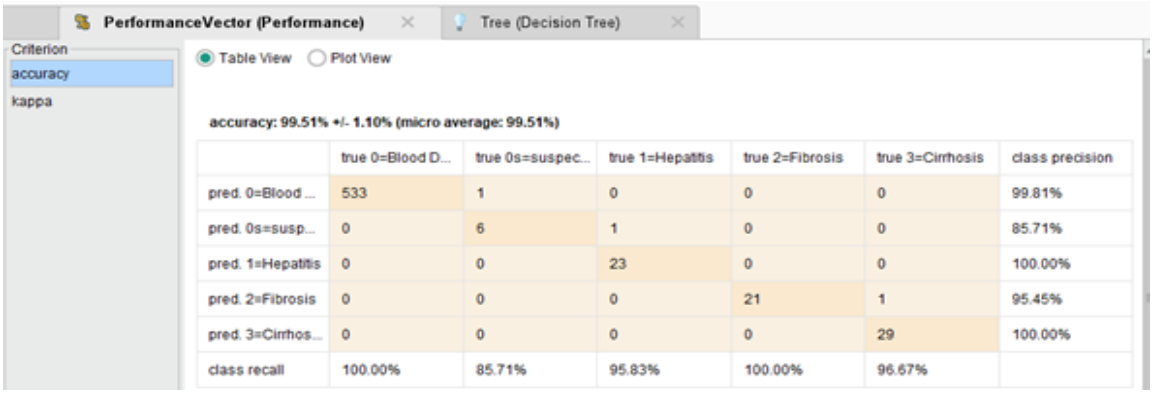
3.2.2 Algoritma Decision Tree

Pengujian model kedua yaitu pengujian menggunakan Algoritma Decision Tree. Berdasarkan tabel Confusion Matrix yang dihasilkan pada rapid miner terdapat pasien yang dinyatakan memiliki riwayat pendonor darah (blood donor) sebanyak 533 orang, pasien yang diduga berpotensi menjadi pendonor darah (suspect blood donor) sebanyak 6 orang, pasien yang menderita penyakit hepatitis sebanyak 23 orang, pasien yang menderita Fibrosis sebanyak 21 orang, sedangkan pasien yang menderita sirosis berjumlah 29 orang. Berikut gambar 8 menunjukkan detail dari tabel *Confusion Matrix* dari hasil pengujian dengan algoritma Decision Tree.

	true 0=Blood D...	true 0s=suspec...	true 1=Hepatitis	true 2=Fibrosis	true 3=Cirrhosis	class precision
pred. 0=Blood ...	533	1	0	0	0	99.81%
pred. 0s=susp...	0	6	1	0	0	85.71%
pred. 1=Hepatitis	0	0	23	0	0	100.00%
pred. 2=Fibrosis	0	0	0	21	1	95.45%
pred. 3=Cirrhos...	0	0	0	0	29	100.00%
class recall	100.00%	85.71%	95.83%	100.00%	96.67%	

Gambar 8. Tabel Confusion Matrix Algoritma Decision Tree

Selain melakukan perhitungan pada confusion matrix, untuk menentukan akurasi Algoritma Decision Tree, dilakukan pengujian menggunakan k-fold cross-validation. Akurasi tersebut diperoleh dari perhitungan 10-fold cross-validation sebesar 99,51%.



	true 0=Blood D...	true 0s=suspec...	true 1=Hepatitis	true 2=Fibrosis	true 3=Cirrhosis	class precision
pred. 0=Blood ...	533	1	0	0	0	99.81%
pred. 0s=susp...	0	6	1	0	0	85.71%
pred. 1=Hepatitis	0	0	23	0	0	100.00%
pred. 2=Fibrosis	0	0	0	21	1	95.45%
pred. 3=Cirrhos...	0	0	0	0	29	100.00%
class recall	100.00%	85.71%	95.83%	100.00%	96.67%	

Gambar 9. Hasil Pengujian Algoritma Decision Tree

3.2.3 Algoritma K-NN + Adaboost

Pengujian ketiga yaitu pengujian menggunakan algoritma K-NN yang dikombinasi dengan Adaboost. Adaboost digunakan untuk meningkatkan performa algoritma K-NN agar mencapai tingkat akurasi yang lebih tinggi daripada penggunaan algoritma K-NN tanpa Adaboost. Berdasarkan tabel Confusion Matrix yang dihasilkan pada rapid miner terdapat pasien yang dinyatakan memiliki riwayat pendonor darah (blood donor) sebanyak 532 orang, pasien yang diduga berpotensi menjadi pendonor darah (suspect blood donor) sebanyak 2 orang, pasien yang menderita penyakit hepatitis sebanyak 18 orang, pasien yang menderita Fibrosis sebanyak 12 orang, sedangkan pasien yang menderita sirosis berjumlah 25 orang. Berikut tabel *Confusion Matrix* dari hasil pengujian menggunakan algoritma K-NN yang dikombinasikan dengan Adaboost.

	true 0=Blood D...	true 0s=suspec...	true 1=Hepatitis	true 2=Fibrosis	true 3=Cirrhosis	class precision
pred. 0=Blood ...	532	5	0	2	3	98.15%
pred. 0s=susp...	0	2	0	0	0	100.00%
pred. 1=Hepatitis	0	0	18	5	1	75.00%
pred. 2=Fibrosis	0	0	2	12	1	80.00%
pred. 3=Cirrhos...	1	0	4	2	25	78.12%
class recall	99.81%	28.57%	75.00%	57.14%	83.33%	

Gambar 10. Tabel Confusion Matrix Algoritma K-NN + Adaboost

Selain melakukan perhitungan pada tabel confusion matrix, untuk menilai akurasi dari algoritma K-NN yang dikombinasikan dengan Adaboost, dilakukan pengujian menggunakan k-fold cross-validation. Hasil perhitungan menggunakan metode *10-fold cross-validation* menunjukkan bahwa akurasi yang dicapai adalah sebesar 95,77%. Angka ini menunjukkan peningkatan akurasi sebesar 2,76% setelah dilakukan optimasi dengan metode Adaboost. Penelitian ini mengindikasikan bahwa penerapan Adaboost mampu meningkatkan kinerja algoritma K-Nearest Neighbor (K-NN) dengan menghasilkan tingkat akurasi yang lebih tinggi dibandingkan tanpa menggunakan Adaboost.

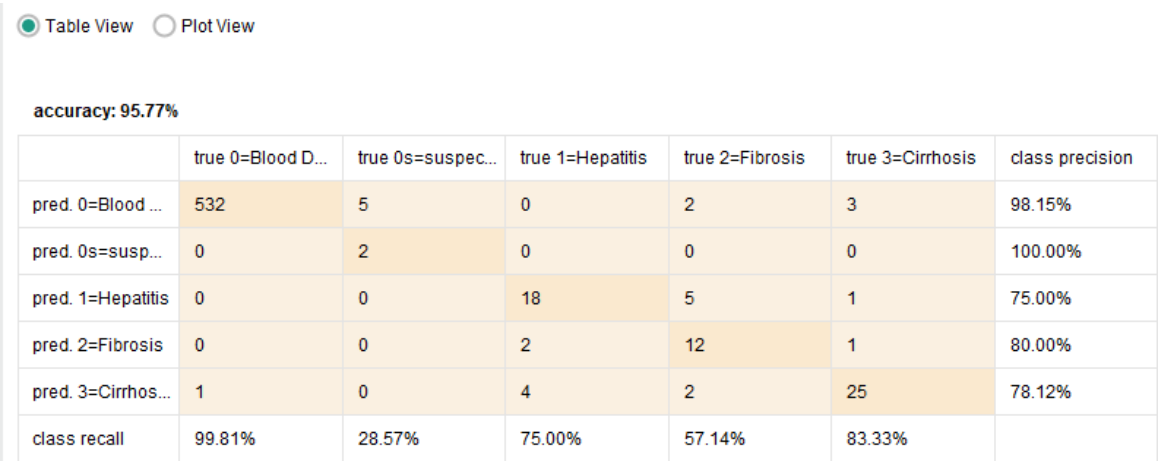


Table View Plot View

accuracy: 95.77%

	true 0=Blood D...	true 0s=suspec...	true 1=Hepatitis	true 2=Fibrosis	true 3=Cirrhosis	class precision
pred. 0=Blood ...	532	5	0	2	3	98.15%
pred. 0s=susp...	0	2	0	0	0	100.00%
pred. 1=Hepatitis	0	0	18	5	1	75.00%
pred. 2=Fibrosis	0	0	2	12	1	80.00%
pred. 3=Cirrhos...	1	0	4	2	25	78.12%
class recall	99.81%	28.57%	75.00%	57.14%	83.33%	

Gambar 11. Pengujian algoritma K-NN + Adaboost

3.2.4 Algortima Decision Tree + Adaboost

Pengujian keempat yaitu pengujian menggunakan algoritma Decision Tree yang dikombinasi dengan Adaboost. Adaboost diterapkan untuk mengoptimalkan kinerja algoritma Decision Tree dengan tujuan memperoleh akurasi yang lebih tinggi dibandingkan dengan penggunaan algoritma Decision Tree tanpa Adaboost. Berdasarkan tabel Confusion Matrix yang dihasilkan pada rapid miner terdapat pasien yang dinyatakan memiliki riwayat pendonor darah (blood donor) sebanyak 533 orang, pasien yang diduga berpotensi menjadi pendonor darah (suspect blood donor) sebanyak 7 orang, pasien yang menderita penyakit hepatitis sebanyak 24 orang, pasien yang menderita Fibrosis sebanyak 21 orang, sedangkan pasien yang menderita sirosis sebanyak 30 orang.

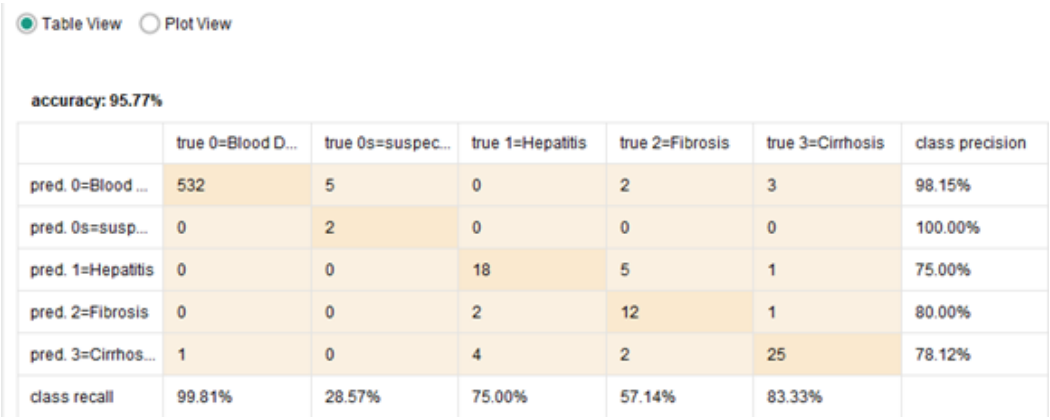


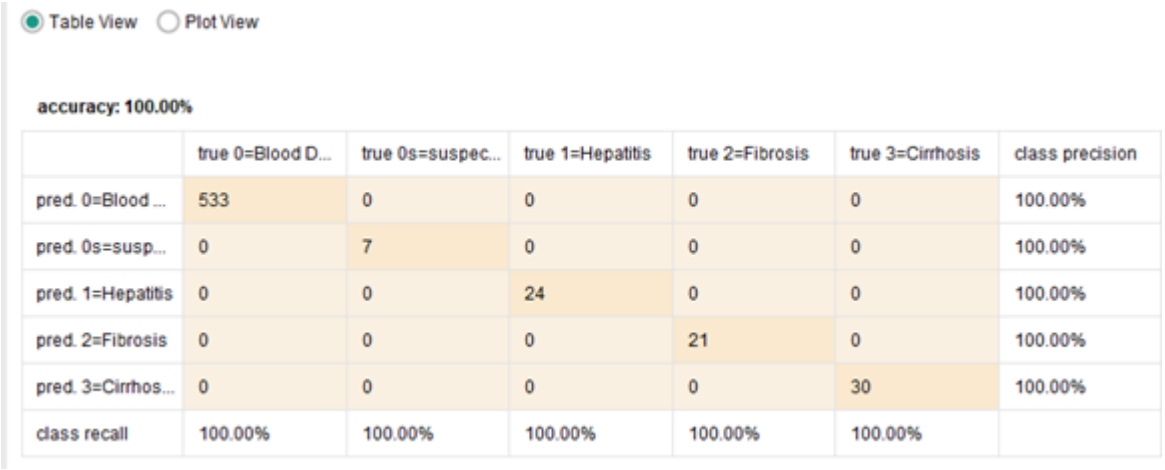
Table View Plot View

accuracy: 95.77%

	true 0=Blood D...	true 0s=suspec...	true 1=Hepatitis	true 2=Fibrosis	true 3=Cirrhosis	class precision
pred. 0=Blood ...	532	5	0	2	3	98.15%
pred. 0s=susp...	0	2	0	0	0	100.00%
pred. 1=Hepatitis	0	0	18	5	1	75.00%
pred. 2=Fibrosis	0	0	2	12	1	80.00%
pred. 3=Cirrhos...	1	0	4	2	25	78.12%
class recall	99.81%	28.57%	75.00%	57.14%	83.33%	

Gambar 12. Tabel Confusion Matrix Algoritma Decision Tree + Adaboost

Selain menghitung performa melalui tabel *Confusion Matrix*, pengujian akurasi algoritma Decision Tree yang dikombinasikan dengan Adaboost juga dilakukan menggunakan metode *k-fold cross-validation*. Hasil perhitungan dengan *10-fold cross-validation* menunjukkan akurasi sebesar 100%. Terdapat peningkatan akurasi sebesar 0,49% setelah dilakukan optimasi menggunakan Adaboost. Temuan ini membuktikan bahwa Adaboost efektif dalam meningkatkan kinerja algoritma Decision Tree dengan menghasilkan akurasi yang lebih tinggi dibandingkan dengan tanpa penerapan Adaboost.


Table View Plot View

accuracy: 100.00%

	true 0=Blood D...	true 0s=suspec...	true 1=Hepatitis	true 2=Fibrosis	true 3=Cirrhosis	class precision
pred. 0=Blood ...	533	0	0	0	0	100.00%
pred. 0s=susp...	0	7	0	0	0	100.00%
pred. 1=Hepatitis	0	0	24	0	0	100.00%
pred. 2=Fibrosis	0	0	0	21	0	100.00%
pred. 3=Cirrhos...	0	0	0	0	30	100.00%
class recall	100.00%	100.00%	100.00%	100.00%	100.00%	

Gambar 13. Hasil Pengujian algoritma Decision Tree + Adaboost

3.3 Analisis Hasil

Berdasarkan empat skenario pengujian algoritma pada dataset penyakit hati, hasil pengujian pertama dengan menggunakan algoritma K-Nearest Neighbor saja menghasilkan akurasi sebesar 93,01%. Pengujian kedua dengan menggunakan algoritma Decision Tree memberikan akurasi sebesar 99,51%. Pada pengujian ketiga, penerapan algoritma K-Nearest Neighbor yang dikombinasikan dengan Adaboost menghasilkan tingkat akurasi sebesar 95,77%. Sedangkan pada pengujian keempat, menggunakan algoritma Decision Tree yang dikombinasikan dengan Adaboost mendapatkan akurasi sebesar 100%. Terdapat peningkatan akurasi sebesar 2,76% dari pengujian ketiga, di mana akurasi awal algoritma K-NN adalah 93,01% dan meningkat menjadi 95,77% setelah diterapkan metode Adaboost. Begitu juga pada pengujian keempat, di mana terdapat peningkatan sebesar 0,49% dari akurasi awal algoritma Decision Tree yang adalah 99,51% dan meningkat menjadi 100% dengan penerapan metode Adaboost. Dengan demikian, dapat disimpulkan bahwa metode Adaboost berhasil meningkatkan dan memperbaiki kinerja dari algoritma K-Nearest Neighbor dan Decision Tree. Dari keempat nilai akurasi yang dihasilkan, penerapan metode Adaboost pada algoritma Decision Tree menunjukkan kinerja tertinggi dengan tingkat akurasi 100%.

4. KESIMPULAN

Penerapan metode Adaboost pada algoritma K-Nearest Neighbor dan Decision Tree dilakukan dengan tujuan untuk meningkatkan kinerja dalam klasifikasi. Penggunaan metode Adaboost memberikan pengaruh yang signifikan terhadap peningkatan nilai akurasi pada algoritma K-Nearest Neighbor dan Decision Tree. Dari hasil evaluasi, algoritma K-Nearest Neighbor menunjukkan akurasi sebesar 93,01%, sementara algoritma Decision Tree mencapai akurasi 99,51%. Setelah penerapan metode Adaboost, akurasi algoritma K-Nearest Neighbor meningkat menjadi 95,77%, dan akurasi algoritma Decision

Tree mencapai 100%. Perbandingan metode Adaboost pada algoritma K-Nearest Neighbor dan Decision Tree membuktikan kemampuannya dalam meningkatkan serta memperbaiki kinerja klasifikasi. Akurasi algoritma K-Nearest Neighbor mengalami peningkatan signifikan sebesar 2,76% setelah dikombinasikan dengan Adaboost, sementara peningkatan akurasi algoritma Decision Tree tercatat sebesar 0,49% setelah penerapan Adaboost. Meskipun peningkatan akurasi pada algoritma K-Nearest Neighbor terbilang cukup mencolok, namun Adaboost juga berhasil memberikan peningkatan yang kurang signifikan, namun tetap terlihat pada algoritma Decision Tree.

5. SARAN

Berdasarkan hasil evaluasi, metode Adaboost terbukti efektif dalam meningkatkan akurasi algoritma klasifikasi, khususnya K-Nearest Neighbor (KNN) dan Decision Tree. Oleh karena itu, disarankan agar metode Adaboost dijadikan salah satu pendekatan standar dalam upaya peningkatan performa model klasifikasi, terutama pada dataset dengan kompleksitas sedang hingga tinggi. Sebagai tindak lanjut dari penelitian ini, disarankan untuk melakukan perbandingan antara metode Adaboost dan teknik ensemble learning lainnya, seperti Bagging dan Gradient Boosting, untuk menentukan metode yang paling efektif sesuai dengan konteks penggunaannya.

DAFTAR PUSTAKA

[1] I. Setiawati, A. P. Wibowo, and A. Hermawan, “Pendahuluan Tinjauan Pustaka Penelitian Sebelumnya Klasifikasi,” *J. Inf. Syst. Manag.*, vol. 1, no. 1, pp. 13–17, 2019.

[2] C. N. Prabiantissa, “Klasifikasi pada Dataset Penyakit HatiMenggunakan Algoritma Support Vector Machine, K-NN, dan Naive Bayes,” *Semin. Nas. Tek. Elektro, Sist. Informasi, dan Tek. Inform.*, vol. 1, no. 1, pp. 263–268, 2021.

[3] Prabiantissa Citra Nurina, “1818-5505-1-Pb,” *Semin. Nas. Tek. Elektro, Sist. Informasi, dan Tek. Inform.*, vol. 1, no. 1, pp. 263–268, 2021.

[4] A. P. Ayudhitama and U. Pujianto, “Analisa 4 Algoritma Dalam Klasifikasi Liver Menggunakan Rapidminer,” *J. Inform. Polinema*, vol. 6, no. 2, pp. 1–9, 2020, doi: 10.33795/jip.v6i2.274.

[5] E. Pusporani, S. Qomariyah, and I. Irhamah, “Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan Machine Learning,” *Inferensi*, vol. 2, no. 1, p. 25, 2019, doi: 10.12962/j27213862.v2i1.6810.

[6] M. D. F. Tino, Herliyani Hasanah, and Tri Djoko Santosa, “Perbandingan Algoritma Support Vector Machines (Svm) Dan Neural Network Untuk Klasifikasi Penyakit Jantung,” *INFOTECH J.*, vol. 9, no. 1, pp. 232–235, 2023, doi: 10.31949/infotech.v9i1.5432.

[7] W. Fadri, “Klasifikasi Penyakit Hati dengan Menggunakan Metode Naive Bayes,” *J. Inf. dan Teknol.*, vol. 5, no. 1, pp. 32–36, 2023, doi: 10.37034/jidt.v5i1.230.

[8] W. Ramdhani, D. Bona, R. B. Musyaffa, and C. Rozikin, “Klasifikasi Penyakit Kangker Payudara Menggunakan Algoritma K-Nearest Neighbor,” *J. Ilm. Wahana Pendidik.*, vol. 8, no. 12, pp. 445–452, 2022.

[9] R. B. Sinaga, D. Widiyanto, B. T. Wahyono, K. Kunci, K. Paru, and R. Forest, “Deteksi Dini Penyakit Kanker Paru dengan Gabungan Algoritma Adaboost dan Random Forest,” pp. 682–691, 2022.

[10] Robianto, S. H. Sitorus, and U. Ristian, “Penerapan Metode Decision Tree Untuk Mengklasifikasikan Mutu Buah Jeruk Berdasarkan Fitur Warna Dan Ukuran,” *J. Komput. dan Apl.*, vol. 09, no. 01, pp. 76–86, 2021.

[11] J. M. A. F. Dina Rachmawaty, "Penerapan Metode Klasifikasi Decision Tree Untuk Memprediksi Kelulusan Tepat Waktu," *J. Ind. Eng. Technol.*, vol. 2, no. 1, pp. 61–74, 2022, doi: 10.24176/jointtech.v2i1.7432.

[12] M. R. Romadhon and F. Kurniawan, "A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia," *3rd 2021 East Indones. Conf. Comput. Inf. Technol. EIConCIT 2021*, pp. 41–44, 2021, doi: 10.1109/EIConCIT50028.2021.9431845.

[13] L. A. Andika, P. A. N. Azizah, and R. Respatiwulan, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019, doi: 10.13057/ijas.v2i1.29998.

[14] L. Swastina, "Penerapan Algoritma C4 . 5 Untuk Penentuan Jurusan Mahasiswa,"

[15] *Gema Aktual.*, vol. 2, no. 1, pp. 93–98, 2018.

[16] A. Byna and M. Basit, "Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 3, pp. 407–411, 2020, doi: 10.32736/sisfokom.v9i3.1023.

[17] L. Pebrianti, F. Aulia, H. Nisa, and K. Saputra S, "Implementation of the Adaboost Method to Optimize the Classification of Diabetes Diseases with the Naïve Bayes Algorithm," *J. Sist. dan Teknol. Inf.*, vol. 7, no. 2, pp. 122–127, 2022, [Online]. Available: http://jurnal.unmuhjember.ac.id/index.php/JUSTINDO