

Komparasi Penerapan Particle Swarm Optimization (PSO) Pada Algoritma Naïve Bayes, K-Nearest Neighbor Dan Random Forest Untuk Prediksi Penyakit Hati

Novian Ridho Saputra, RZ. Abdul Aziz*

Magister Teknik Informatika, Fakultas Komputer
IIB Darmajaya
Bandar Lampung, Indonesia
rz_aziz@darmajaya.ac.id

Abstract- The liver plays a very important role as the metabolic center, functioning to meet the needs of the brain and as a blood filter for harmful substances coming from the intestines. Impaired liver function can lead to various diseases such as liver disease, hepatitis, liver cancer, liver cirrhosis, and other liver-related conditions. The liver is also vulnerable to disease, which can result in its inability to function normally, potentially leading to death. Liver disease is a condition that has long existed and is quite common in society. According to data from WHO (World Health Organization), nearly 1.2 million people per year, especially in Southeast Asia and Africa, die from liver disease. In this research, the Naïve Bayes, K-NN, Random Forest, and PSO algorithms will be applied by optimizing the attributes derived from the dataset to predict liver disease. The accuracy levels achieved in the comparison of the three algorithms, namely Random Forest and PSO, is 99.84%, PSO and K-NN is 98.54%, while PSO and Naïve Bayes achieve an accuracy of 94.65%. From this, it can be concluded that the addition of PSO to Random Forest can enhance the accuracy level to a better degree.

Keywords: Naïve Bayes, K-Nearest Neighbor, Random Forest, PSO, Liver Disease.

Abstrak- Hati memiliki peran yang sangat penting, yaitu sebagai pusat metabolisme yang berfungsi untuk memenuhi kebutuhan otak dan sebagai penyaring darah dari zat-zat berbahaya yang berasal dari usus. Gangguan fungsi hati dapat menyebabkan berbagai penyakit seperti penyakit hati, hepatitis, kanker hati, sirosis hati, serta penyakit hati lainnya. Hati juga rentan terhadap serangan penyakit yang dapat mengakibatkan ketidakmampuan hati untuk berfungsi dengan normal, bahkan berpotensi menyebabkan kematian. Penyakit hati adalah kondisi yang telah lama ada dan cukup umum di masyarakat. Menurut data dari WHO (Organisasi Kesehatan Dunia), hampir 1,2 juta orang setiap tahunnya, terutama di Asia Tenggara dan Afrika, meninggal akibat penyakit hati. Dalam penelitian ini, algoritma Naïve Bayes, K-NN, Random Forest, dan PSO akan diterapkan dengan mengoptimalkan atribut-atribut yang diambil dari dataset untuk memprediksi penyakit hati. Tingkat akurasi yang dihasilkan dalam perbandingan tiga algoritma, yaitu Random Forest dan PSO, adalah 99,84%, PSO dan K-NN mencapai akurasi 98,54%, sedangkan PSO dan Naïve Bayes menghasilkan akurasi 94,65%. Dari sini, dapat disimpulkan bahwa penambahan PSO pada Random Forest dapat meningkatkan tingkat akurasi ke level yang lebih baik.

Kata Kunci: Naïve Bayes, K - Nearest Neighbor, Random Forest, PSO, Penyakit Hati

1. Pendahuluan

Dunia kesehatan mengumpulkan banyak data tentang kesehatan, tetapi masih sulit untuk mendapatkan Keakuratan dalam prediksi penyakit dalam bidang kesehatan sangat penting karena keputusan yang efektif diperlukan. *One of the most significant aspects of human life is health, which serves as the foundation for many scientific discoveries that result in medications, medical equipment, or advancements in the health sector [1], The liver, the biggest organ in the body, is*

necessary for both food digestion and the body's release of toxins. Alcohol use combined with viruses can cause liver damage and put a person's health in danger. Liver illnesses include cirrhosis, hepatitis, liver tumors, liver cancer, and many other conditions. Among them, cirrhosis and liver disorders are the leading causes of mortality [2]. Tobacco use, inactivity, excessive alcohol use, and a poor diet all raise the chance of dying from chronic illnesses [3]. Salah satu organ tubuh kita yang sangat besar dan penting



adalah hati. Hepatosit adalah sel epitel dengan bentuk khusus yang merupakan bagian penting dari hati ini. Hati pada dasarnya adalah kelenjar eksorin karena mensekresi cairan empedu ke *duodenum*. Ini juga merupakan kelenjar *endokrin* dan penyaring darah [4]. Penyakit hati yang sudah akut memengaruhi fungsi hati. Gejala penyakit hati dapat diidentifikasi melalui gejala klinis dan fisik pasien; gejala klinis dapat diidentifikasi melalui perasaan pasien dan kondisi fisik pasien, gejala penyakit hati sangat beragam dan kompleks, dan gejalanya mirip dengan beberapa penyakit lain. Hati adalah pusat metabolisme, yang menjaga kebutuhan otak dan membersihkan darah dari zat-zat berbahaya dari usus. Kelainan fungsi hati ini dapat menyebabkan berbagai penyakit hati, termasuk liver, hepatitis, kanker hati, sirosis hati, dan penyakit hati lainnya. Penyakit hati kompleks seperti kanker hati membutuhkan screening pasien yang tepat dan akurat. Sekitar 0,6% dari 2,9 juta orang yang tinggal di Indonesia mengidap hepatitis. Kementerian Kesehatan telah berusaha untuk mencegah infeksi HB 0–4 sejak bayi. Data dari Kementerian Kesehatan menunjukkan bahwa jumlah orang yang diimunisasi telah melampaui target, tetapi ini belum dapat mengubah jumlah orang yang menderita hepatitis di Indonesia [5].

Penyakit liver adalah penyakit hati yang sudah lama ada dan cukup umum di masyarakat. Menurut data *WHO (World Health Organization)*, penyakit liver menyebabkan hampir 1,2 juta kematian setiap tahun, terutama di Asia Tenggara dan Afrika. Faktor penyebab penyakit liver diantaranya kelainan hati yang sudah ada sejak lama dan dapat menyebabkan hati tidak mampu berfungsi dengan baik atau bahkan menyebabkan kematian [6].

Berdasarkan masalah di atas, maka peneliti dalam hal ini mengambil judul “Komparasi Penerapan Particle Swarm Optimization (PSO) Pada Algoritma Naïve Bayes, K-Nearest Neighbor Dan Random Forst Untuk Prediksi Penyakit Hati”. Dalam penelitian ini akan dilakukan penerapan algoritma Naïve Bayes, K-NN, Random Forest dan PSO dengan mengoptimalkan atribut-atribut yang berasal dari Dataset untuk memprediksi penyakit Hati, dan tools yang digunakan adalah rapid miner sehingga dapat mengetahui performa yang baik dari algoritma tersebut untuk mengetahui penyakit Hati.

2. Metodologi

A. Data

Dataset yang akan digunakan dalam penelitian ini berasal dari Kaggle, khususnya dari tautan berikut: <https://www.kaggle.com/code/lutfiandri/klasifikasi-penyakit-hati-liver/input>. Dataset ini terdiri dari 615 data dan akan dianalisis menggunakan alat Rapid Miner. Terdapat 13 atribut yang dianalisis dalam dataset ini, yaitu: *Category*, *Age*, *Sex*, *ALB (Albumin)*, *ALP (Alkaline Phosphatase)*, *ALT (Alanine Aminotransferase)*, *AST (Aspartate Aminotransferase)*, *BIL (Bilirubin)*, *CHE (Cholinesterase)*, *CHOL (Cholesterol)*, *CREA (Creatinine)*, *GGT (Gamma-Glutamyl Transferase)*, dan *PROT (Protein)*.

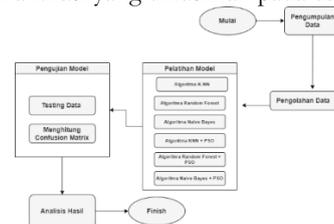
B. Alur Penelitian

Langkah awal dalam penelitian ini adalah pengumpulan data. Data yang digunakan bersumber dari dataset publik yang tersedia di Kaggle.

Pada tahap berikutnya, fokus utama adalah pengelolaan data yang telah dikumpulkan. Langkah ini bertujuan untuk memastikan bahwa data dapat diolah secara efisien, relevan, dan aman. Dataset yang digunakan diambil dari Kaggle dengan jumlah 615 data, tersedia di <https://www.kaggle.com/code/lutfiandri/klasifikasi-penyakit-hati-liver/input>. Pengolahan data dilakukan menggunakan alat RapidMiner dengan Terdapat 13 atribut yang dianalisis dalam dataset ini, yaitu: *Category*, *Age*, *Sex*, *ALB (Albumin)*, *ALP (Alkaline Phosphatase)*, *ALT (Alanine Aminotransferase)*, *AST (Aspartate Aminotransferase)*, *BIL (Bilirubin)*, *CHE (Cholinesterase)*, *CHOL (Cholesterol)*, *CREA (Creatinine)*, *GGT (Gamma-Glutamyl Transferase)*, dan *PROT (Protein)*. Metode yang diterapkan dalam penelitian ini meliputi *PSO*, *Naïve Bayes*, *K-NN*, dan *Random Forest*.

Tahap ketiga adalah penerapan dan pengujian metode yang digunakan untuk memperoleh hasil terbaik. *PSO* diterapkan sebagai metode untuk meningkatkan akurasi dari hasil yang diperoleh.

Tahap terakhir adalah analisis hasil, di mana tahap ini mengevaluasi keberhasilan model yang digunakan berdasarkan akurasi yang dihasilkan pada data uji



Gambar 1. Alur Dalam Tahapan Penelitian

2.1 Penyakit Hati

Hati merupakan salah satu organ vital dalam tubuh manusia yang memiliki kemampuan luar biasa dalam menjalankan berbagai fungsi penting, termasuk metabolisme nutrisi, detoksifikasi racun, dan produksi empedu. Namun, hati rentan terhadap berbagai penyakit yang dapat mempengaruhi kinerjanya, yang berkembang melalui beberapa stadium, mulai dari peradangan (hepatitis), fibrosis, sirosis, hingga gagal hati. Penyakit hati dapat dikenali melalui gejala klinis dan fisik seperti kelelahan ekstrem, kulit dan mata menguning (*jaundice*), serta pembengkakan perut. Peningkatan kasus penyakit hati disebabkan oleh berbagai faktor, terutama gaya hidup tidak sehat yang tidak terkontrol. Faktor-faktor risiko utama termasuk diabetes melitus, dislipidemia, dan hipertensi. Pengelolaan gaya hidup yang lebih sehat dan deteksi dini gejala dapat membantu mencegah perkembangan penyakit hati lebih lanjut [7].

2.2 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) merupakan metode optimasi yang sangat mudah diterapkan dan memungkinkan modifikasi berbagai parameter. Dalam

PSO, beberapa teknik yang digunakan meliputi peningkatan bobot atribut untuk setiap variabel atau atribut yang digunakan, pemilihan atribut (*attribute selection*), serta seleksi atribu [8]. Algoritma Particle Swarm Optimization (PSO) dikembangkan dengan mengamati perilaku sosial hewan seperti burung, lebah, dan ikan. Dalam algoritma ini, setiap hewan dianggap sebagai partikel, di mana kecerdasan individu partikel serta kecerdasan kolektif dalam kelompoknya mempengaruhi pergerakan masing-masing partikel. Ketika satu partikel menemukan jalur tercepat dan paling efisien menuju sumber makanan, partikel lain akan mengikuti. PSO dapat diterapkan dalam penelitian penyakit hati untuk berbagai keperluan, seperti optimalisasi pemilihan fitur, peningkatan akurasi model prediksi, dan pemodelan diagnosis penyakit. Sebagai contoh, dalam memprediksi perkembangan penyakit hati berdasarkan data medis pasien, PSO dapat digunakan untuk memilih fitur-fitur yang paling relevan dari dataset yang besar. Proses ini melibatkan penempatan partikel-partikel di ruang pencarian fitur, di mana setiap partikel mewakili satu set fitur terpilih. Selama iterasi, partikel-partikel tersebut akan menyesuaikan posisi dan kecepatannya untuk menemukan kombinasi fitur terbaik berdasarkan evaluasi model prediksi (seperti tingkat akurasi atau kesalahan prediksi). Selain itu, PSO juga dapat dimanfaatkan untuk mengoptimalkan parameter dalam model prediksi yang kompleks. Dalam konteks ini, setiap partikel mewakili satu set parameter model. Melalui iterasi, partikel-partikel tersebut akan memperbarui posisi dan kecepatannya berdasarkan kinerja model, dengan tujuan menemukan parameter optimal yang menghasilkan prediksi terbaik untuk kondisi penyakit hati. [9].

2.3 Naïve Bayes

Algoritma Naïve Bayes menggunakan metode percabangan dalam matematika untuk menentukan kemungkinan klasifikasi yang paling tepat berdasarkan frekuensi klasifikasi pada data pelatihan. Pendekatan ini dikenal sebagai teori probabilitas. Rumus perhitungan Naïve Bayes adalah sebagai berikut [10].

$$P(X|Y) = \frac{P(Y|X) \times P(X)}{P(Y)}$$

Keterangan :

- Y = data dengan kelas yang belum diketahui
- X = hipotesis data Y merupakan suatu kelas spesifik
- P(X | Y) = probabilitas hipotesis X berdasarkan kondisi Y
- P(X) = Probabilitas hipotesis
- P(Y | X) = Probabilitas Y berdasarkan kondisi pada hipotesis
- P(Y) = Probabilitas Y

Dalam teori Bayes, terdapat beberapa prinsip mendasar yang menekankan ketidaktergantungan setiap fitur terhadap data lain yang tidak terkait pada kumpulan data yang sama. Hipotesis Naïve Bayes merupakan label identitas kelas yang digunakan dalam pemetaan klasifikasi, yang juga berhubungan dengan korelasi hipotesis. Fitur-fitur yang dimasukkan ke dalam model klasifikasi dapat dianggap sebagai bukti [9]. Selain itu, Algoritma Naïve

Bayes Classifier memiliki performa yang sangat baik dalam beberapa kasus klasifikasi teks.

2.4 K – Nearest Neighbor

Selain mudah digunakan, algoritma KNN banyak digunakan untuk klasifikasi. Algoritma ini memiliki kelemahan dalam menentukan nilai k; jika nilai k terlalu besar atau k=1, hasil klasifikasi akan menjadi tidak jelas atau kabur, sedangkan jika nilai k terlalu kecil atau k=1, hasil klasifikasi akan terasa kaku karena tidak ada pilihan. Oleh karena itu, penelitian untuk menentukan nilai k yang tepat diperlukan [11]. Algoritma *K-Nearest Neighbor (K-NN)* adalah metode untuk mengklasifikasikan objek baru berdasarkan huruf K tetangga terdekatnya. *K-NN* termasuk dalam algoritma pengajaran yang diawasi, di mana hasil dari pertanyaan instance yang baru dikategorikan menurut mayoritas kategori *K-NN*. Hasil klasifikasi ini adalah kelas yang paling sering muncul [12]. Untuk mengukur jarak terdekat antara data pelatihan, kedua jarak geometri dan jarak *Manhattan* akan digunakan. Penelitian sebelumnya menunjukkan bahwa jarak geometri lebih akurat daripada jarak *Manhattan* [13]. Berikut adalah persamaan Euclidean Distance dalam Algoritma K-NN :

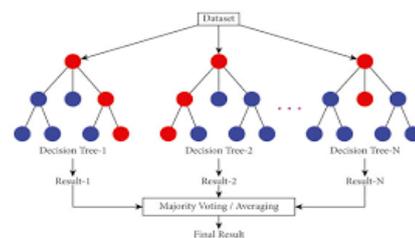
$$k_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

Keterangan :

- x_1 = Data Uji
- x_2 = Data Training
- i = Variabel Data
- p = Dimensi Data
- k = Jarak
- p = Dimensi Data

2.5 Random Forest

Random Forest merupakan teknik klasifikasi yang memanfaatkan konsep *Decision Tree*. Teknik ini menggunakan pemilihan atribut secara acak pada setiap node untuk menentukan hasil klasifikasi. Selama proses klasifikasi, keputusan akhir diperoleh dari suara terbanyak yang dihasilkan oleh pohon keputusan. Hutan acak dibentuk melalui metode bagging dengan atribut yang dipilih secara acak. Pohon keputusan dalam *Random Forest* tidak mengalami pengurangan, melainkan terus berkembang hingga mencapai kondisi optimal. Akhirnya, terbentuklah hutan yang terdiri dari kumpulan pohon keputusan terstruktur dengan vektor acak independen yang didistribusikan secara identik. *Random Forest* kemudian memberikan suara pada kelas yang paling dominan berdasarkan input x [14].



Gambar 2. Model Algoritma Random Forest

3. Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini adalah dataset dari Kaggle yaitu data penyakit hati , Jumlah record data 615 data terdiri dari 12 atribut dan 1 Label. Atribut yang digunakan *Age, Sex, ALB (Albumin), ALP(Alkaline Phosphatase), ALT (Alanine Aminotransferase),*

AST (Aspartate Aminotransferase), BIL (Bilirubin) , CHE (Cholinesterase), CHOL (Cholesterol), CREA (Creatinine), GGT (Gamma - Glutamyl Transferase), PROT (Protein). Kumpulan data ini digunakan untuk memprediksi apakah seorang pasien kemungkinan besar menderita penyakit hati.

A. Penelitian Menggunakan Algoritma Naïve Bayes

Pada tabel dibawah ini, data yang telah disiapkan diimplementasikan pada aplikasi *Rapidminer* untuk melakukan prediksi penyakit hati. Proses ini melibatkan eksperimen menggunakan metode *cross validation*. *Cross validation* merupakan sebuah teknik yang umum digunakan dalam evaluasi model di bidang *data mining* dan *machine learning*. Teknik ini secara langsung membagi *dataset* menjadi dua bagian utama, yaitu *data training* dan *data testing*. Dalam penelitian ini, data yang digunakan adalah *data supervised*, yang berarti kita memiliki label atau target yang ingin diprediksi. Algoritma yang dipilih untuk prediksi adalah Naïve Bayes. Tujuan dari *cross validation* adalah untuk mengukur kinerja model secara keseluruhan dengan mempertimbangkan variasi yang mungkin terjadi dalam pembagian data. Hasil dari eksperimen tersebut dapat di lihat pada tabel di bawah ini.

Tabel 1 Confusion Matrix Hasil Prediksi Penyakit Hati Menggunakan Algoritma Naïve Bayes pada Rapidminer

Accuracy : 98.06%						
	true 0=blood donor	true 0s=suspect blood	True1= Hepatitis	True 2=Fibrosis	True 3=Cirrhosis	Class Precision
Pred 0=blood donor	532	0	0	0	1	99.8 1%
Pred 0s=suspect blood	0	7	0	0	0	100. 00%
Pred 1=Hepatitis	1	0	20	3	0	83.3 3%
Pred 2=Fibrosis	0	0	2	17	2	80.9 5%
Pred 3=Cirrhosis	0	0	2	1	27	90.0 0%
Class recall	99.8 1%	100. 00%	83.33%	80.95 %	90.00 %	

- [1] Prediksi Kelas:
- [2] Angka dalam sel menunjukkan jumlah kejadian.

- [3] True 0 (Pendonor Darah):
- [4] Diprediksi sebagai Pendonor Darah: 532
- [5] Diprediksi sebagai Sirosis: 1
- [6] True 0s (Darah Terduga):
- [7] Diprediksi sebagai Darah Terduga: 7
- [8] True 1 (Hepatitis):
- [9] Diprediksi sebagai Hepatitis: 20
- [10] Diprediksi sebagai Pendonor Darah: 1
- [11] Diprediksi sebagai Sirosis: 3
- [12] True 2 (Fibrosis):
- [13] Diprediksi sebagai Hepatitis: 2
- [14] Diprediksi sebagai Fibrosis: 17
- [15] Diprediksi sebagai Sirosis: 2
- [16] True 3 (Sirosis):
- [17] Diprediksi sebagai Sirosis: 27
- [18] Diprediksi sebagai Fibrosis: 1

B. Penelitian Menggunakan Algoritma K - NN

Hasil Penerapan data pada Rapidminer untuk Prediksi Penyakit Hati menggunakan algoritma K-nn ditunjukkan pada tabel 2 dibawah ini:

Tabel 2 Confusion Matrix Hasil Prediksi Penyakit Hati Menggunakan Algoritma K-NN pada Rapidminer

Accuracy : 92.53%						
	true 0=blood donor	true 0s=suspect blood	True1= Hepatitis	True 2=Fibrosis	True 3=Cirrhosis	Class Precision
Pred 0=blood donor	530	5	1	2	5	97.6 1%
Pred 0s=suspect blood	0	1	0	0	0	100. 00%
Pred 1=Hepatitis	0	1	13	8	3	52.0 0%
Pred 2=Fibrosis	2	0	5	9	6	40.9 1%
Pred 3=Cirrhosis	1	0	5	2	16	66.6 7%
Class recall	99.4 4%	14. 29 %	54.17%	42.86 %	53.33 %	



Prediksi Kelas:
Angka-angka dalam sel menunjukkan jumlah kejadian.

- [19] True 0 (Pendonor Darah):
- [20] Diprediksi sebagai Pendonor Darah: 530
- [21] Diprediksi sebagai Darah Terduga: 5
- [22] Diprediksi sebagai Hepatitis: 1
- [23] Diprediksi sebagai Fibrosis: 2
- [24] Diprediksi sebagai Sirosis: 5
- [25] True 0s (Darah Terduga):
- [26] Diprediksi sebagai Darah Terduga: 1
- [27] True 1 (Hepatitis):
- [28] Diprediksi sebagai Hepatitis: 13
- [29] Diprediksi sebagai Fibrosis: 8
- [30] True 2 (Fibrosis):
- [31] Diprediksi sebagai Hepatitis: 1
- [32] Diprediksi sebagai Fibrosis: 5
- [33] Diprediksi sebagai Sirosis: 9
- [34] True 3 (Sirosis):
- [35] Diprediksi sebagai Sirosis: 8
- [36] Diprediksi sebagai Fibrosis: 1

C. Penelitian Menggunakan Algoritma Random Forest

Hasil Penerapan data pada Rapidminer untuk Prediksi Penyakit Hati menggunakan algoritma Random Forest ditunjukkan pada tabel 3 dibawah ini:

Tabel 3 Confusion Matrix Hasil Prediksi Penyakit Hati Menggunakan Algoritma Random Forest pada Rapidminer

Accuracy : 99.51%

	true 0=blood donor	true 0s=suspect blood	True1=Hepatitis	True 2=Fi brosis	True 3=Cir thosis	Class Precision
Pred 0=blood donor	533	1	0	0	0	99.81%
Pred 0s=suspect blood	0	6	0	0	0	100.00%
Pred 1=Hepatitis	0	0	24	1	0	96.00%
Pred 2=Fi brosis	0	0	0	20	1	95.24%
Pred 3=Cir thosis	0	0	0	0	29	100.00%
Class recall	100.00%	85.71%	100.00%	95.24%	96.67%	

Prediksi Kelas:
[1] Angka dalam sel menunjukkan jumlah kasus.

- [2] True 0 (Pendonor Darah):
- [3] Diprediksi sebagai Pendonor Darah: 533
- [4] Diprediksi sebagai Darah Terduga: 1
- [5] True 0s (Darah Terduga):
- [6] Diprediksi sebagai Darah Terduga: 6
- [7] True 1 (Hepatitis):
- [8] Diprediksi sebagai Hepatitis: 24
- [9] True 2 (Fibrosis):
- [10] Diprediksi sebagai Fibrosis: 20
- [11] True 3 (Sirosis):
- [12] Diprediksi sebagai Sirosis: 29

D. Penelitian Menggunakan Algoritma Naïve Bayes dan PSO

Pada tabel didalam proses penelitian ini, data yang telah diolah terlebih dahulu diterapkan pada aplikasi *Rapidminer*, suatu *platform* yang digunakan untuk melakukan analisis data secara efisien. Penggunaan data tersebut bertujuan untuk memprediksi kemungkinan

Tabel 4 Confusion Matrix PSO Dan Naïve Bayes

Accuracy : 94.65%

	true 0=blood donor	true 0s=suspect blood	True1=Hepatitis	True 2=Fi brosis	True 3=Cir thosis	Class Precision
Pred 0=blood donor	532	5	0	3	5	97.61%
Pred 0s=suspect blood	0	1	0	0	0	100.00%
Pred 1=Hepatitis	0	0	16	5	0	76.19%
Pred 2=Fi brosis	0	0	3	11	3	64.71%
Pred 3=Cir thosis	1	1	5	2	22	70.97%
Class recall	99.81%	14.29%	66.67%	52.38%	73.33%	

penyakit hati dengan menerapkan teknik *cross validation*, sebuah metode yang digunakan untuk memastikan bahwa model yang dibangun dapat diterapkan dengan baik pada data yang baru. Dalam eksperimen ini, tidak hanya satu, tetapi dua algoritma digunakan, yakni *PSO* dan *Naïve Bayes*. Hasil eksperimen yang dihasilkan dari proses ini akan ditampilkan dalam tabel 4, yang akan memberikan gambaran tentang seberapa baik model yang telah dibangun dalam memprediksi penyakit hati.

Tabel 4 adalah *confusion matrix* yang menunjukkan hasil eksperimen pada sebuah gambar atau tabel yang menampilkan *confusion matrix*, yang merupakan sebuah alat visual untuk mengevaluasi kinerja model klasifikasi. *Confusion matrix* tersebut berisi hasil dari eksperimen yang



dilakukan dengan menggunakan teknik *PSO (Particle Swarm Optimization)* dan *Naive Bayes*. Dalam *confusion matrix* kita dapat melihat *PSO* dan *Naive Bayes* hasil akurasi yang dihasilkan adalah 94.65%.

E. Penelitian Menggunakan Algoritma K – NN dan PSO

Hasil Penerapan *PSO* dan algoritma *K – NN* pada *rapidminer* ditunjukkan pada tabel 5 yang ada di bawah ini :

Tabel 5 Confusion Matrix *PSO* Dan *K – NN*

Accuracy : 98.54%						
	true 0=blood donor	true 0s=suspect blood	True1= Hepatitis	True 2=Fi brosis	True 3=Cir thosis	Clas s Prec ision
Pred 0=blo od donor	532	0	0	0	1	99.8 1%
Pred 0s=su spect blood	1	7	0	0	0	87.5 0%
Pred 1=He patitis	0	0	21	2	0	91.3 0%
Pred 2=Fi brosis	0	0	2	19	2	82.6 1%
Pred 3=Cir thosis	0	0	1	0	27	96.4 3%
Class recall	99.8 1%	100. 00%	87.50%	90.48 %	90.00 %	

Tabel 5 adalah *confusion matrix* yang menunjukkan hasil eksperimen pada sebuah gambar atau tabel yang menampilkan *confusion matrix*, yang merupakan sebuah alat visual untuk mengevaluasi kinerja model klasifikasi. *Confusion matrix* tersebut berisi hasil dari eksperimen yang dilakukan dengan menggunakan teknik *PSO (Particle Swarm Optimization)* dan *K – NN*. Dalam *confusion matrix* kita dapat melihat *PSO* dan *K – NN* hasil akurasi yang dihasilkan adalah 98.54%.

F. Penelitian Menggunakan Algoritma Random Forest dan PSO

Hasil Penerapan *PSO* dan algoritma *Random Forest* pada *rapidminer* ditunjukkan pada Tabel 6 yang ada di bawah ini:

Tabel 6 Confusion Matrix *PSO* Dan *Random Forest*

Accuracy : 99.84%						
	true 0=blood donor	true 0s=suspect blood	True1= Hepatitis	True 2=Fi brosis	True 3=Cir thosis	Clas s Prec ision
Pred 0=blo od donor	533	0	0	0	0	100. 00%

Pred 0s=su spect blood	0	7	0	0	0	100. 00%
Pred 1=He patitis	0	0	24	1	0	96.0 0%
Pred 2=Fi brosis	0	0	0	20	0	100. 00%
Pred 3=Cir thosis	0	0	0	0	30	100. 00%
Class recall	100. 00%	100. 00%	100.00%	95.24 %	100.0 0%	

Tabel 6 adalah *confusion matrix* yang menunjukkan hasil eksperimen pada sebuah gambar atau tabel yang menampilkan *confusion matrix*, yang merupakan sebuah alat visual untuk mengevaluasi kinerja model klasifikasi. *Confusion matrix* tersebut berisi hasil dari eksperimen yang dilakukan dengan menggunakan teknik *PSO (Particle Swarm Optimization)* dan *Random Forest*. Dalam *confusion matrix* kita dapat melihat *PSO* dan *Random Forest* hasil akurasi yang dihasilkan adalah 99.84%.

G. Hasil Perbandingan Metode

Berikut ini merupakan tabel hasil akurasi dari *Naive Bayes*, *K – NN* dan *Random Forest* yang menggunakan metode *PSO* dan tanpa menggunakan *PSO*:

Tabel 7 Hasil Perbandingan Akurasi Algoritma Klasifikasi

Metode	Hasil Accuracy
Naive Bayes	98,06 %
K – NN	92,53 %
Random Forest	99,51 %
Naive Bayes + PSO	94,65 %
K – NN + PSO	98,54 %
Random Forest + PSO	99,84 %
Hasil Metode Terbaik:	99,84 %
Random Forest + PSO	

Dari tabel 7 dapat dilihat hasil akurasi tanpa Menggunakan *PSO*, *Naive Bayes* mendapatkan hasil 98,06 %, *K – NN* mendapatkan hasil akurasi sebesar 92,55 %, *Random Forest* mendapatkan hasil akurasi sebesar 99,51 %. Sedangkan algoritma klasifikasi yang menggunakan metode *PSO* terlihat perbedaan yaitu *Naive Bayes* yang menggunakan *PSO* mendapatkan hasil akurasi sebesar 94,65 %, *K – NN* dengan menggunakan *PSO* mendapatkan hasil akurasi sebesar 98,54 %, dan *Random Forest* setelah menggunakan *PSO* mendapatkan hasil akurasi sebesar 99,84 %.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa penggunaan kombinasi metode *PSO* dan algoritma *Random Forest* dalam prediksi penyakit hati menghasilkan tingkat akurasi yang lebih baik dibandingkan dengan menggunakan algoritma *Naive Bayes* dan *K – NN*. Hasil analisis menunjukkan bahwa tingkat akurasi dari penelitian ini mencapai 99.84%. Dengan



menggunakan kombinasi *Random Forest* dan *PSO*, mengindikasikan kemampuan model untuk mengidentifikasi kasus penyakit hati secara tepat. Dengan demikian, dapat disimpulkan bahwa kombinasi antara *Random Forest* dan *PSO* adalah pilihan yang efektif dalam meningkatkan akurasi prediksi penyakit hati dalam konteks penelitian ini.

5. Daftar Pustaka

- [1] I. Herliawan *et al.*, “Classification of Liver Disease By Applying Random Forest Algorith and Backward Elimination,” *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komput.*, vol. 6, no. 1, pp. 89–94, 2020, doi: 10.33480/jitk.v6i1.1424.
- [2] A. K. M. S. Rahman, F. M. Javed Mehedi Shamrat, Z. Tasnim, J. Roy, and S. A. Hossain, “A comparative study on liver disease prediction using supervised machine learning algorithms,” *Int. J. Sci. Technol. Res.*, vol. 8, no. 11, pp. 419–422, 2019.
- [3] A. Siburian and C. F. Chang, “Chronic Liver Disease Lowering Physical and Mental Health Dimensions,” *J. Keperawatan Indones.*, vol. 24, no. 3, pp. 149–156, 2021, doi: 10.7454/JKI.V24I3.1240.
- [4] M. Kesuma, “Prediksi Penyakit Liver Menggunakan Algoritma Random Forest,” *J. Inf. dan Komput.*, vol. 11, no. 2, p. 2023, 2023.
- [5] T. Indriyani, G. E. M. Kurniawan, Yuliasuti, A. Rachman, C. N. Prabiantissa, and R. K. Hapsari, “An Improve KNN Method for Classification of Sexually Transmitted Diseases,” *Sixth Int. Conf. Vocat. Educ. Electr. Eng.*, 2023.
- [6] A. P. Ayudhitama and Utomo 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.
- [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] A. Mustopa, Hermanto, Anna, E. B. Pratama, A. Hendini, and D. Risdiansyah, “Analysis of user reviews for the pedulilindungi application on google play using the support vector machine and naive bayes algorithm based on particle swarm optimization,” *2020 5th Int. Conf. Informatics Comput. ICIC 2020*, vol. 19, 2020, doi: 10.1109/ICIC50835.2020.9288655.
- [9] S. H. F. Hakim, I. Cholissodin, and A. W. Widodo, “Seleksi Fitur Dengan Particle Swarm Optimization Untuk Pengenalan Pola Wajah Menggunakan Naive Bayes (Studi Kasus Pada Mahasiswa Universitas Brawijaya Fakultas Ilmu Komputer Gedung A),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 10, pp. 1045–1057, 2017.
- [10] R. Dahlia, S. Hadiani, R. Dahlia, N. Wuryani, W. Gata, and A. Selawati, “Penerapan Data Mining Terhadap Data Covid-19 Menggunakan Algoritma Klasifikasi,” *Sekol. Tinggi Ilmu Komput. Nusa Mandiri Jakarta*, vol. 21, no. 1, pp. 44–52, 2021.
- [11] Widhi Ramdhani, D. Bona, R. B. Musyaffa, and C. Rozikin, “Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor,” vol. 8, no. 12, pp. 1–23, 2022.
- [12] M. Syukri Mustafa and I. Wayan Simpen, “Implementasi Algoritma K-Nearest Neighbor (KNN) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyampa Kabupaten Bulukumba,” *Februari*, vol. 2019, no. 1, pp. 1–10, 2019.
- [13] N. Khairina, T. T. S. Sibarani, R. Muliono, Z. Sembiring, and M. Muhathir, “Identification of Pneumonia using The K-Nearest Neighbors Method using HOG Fitur Feature Extraction,” *J. Informatics Telecommun. Eng.*, vol. 5, no. 2, pp. 562–568, 2022, doi: 10.31289/jite.v5i2.6216.
- [14] L. Ratnawati and D. R. Sulistyaningrum, “Penerapan Random Forest untuk Mengukur Tingkat Keperahan Penyakit pada Daun Apel,” *J. Sains dan Seni ITS*, vol. 8, no. 2, 2020, doi: 10.12962/j23373520.v8i2.48517.

