

## CESS

(Journal of Computer Engineering, System and Science)

Available online: <https://jurnal.unimed.ac.id/2012/index.php/cess>

ISSN: 2502-714x (Print) | ISSN: 2502-7131 (Online)



### Implementasi Algoritma *Random Forest* dengan Variasi Parameter $n\_estimators$ untuk Klasifikasi Penyakit Hati

### *Implementation of the Random Forest Algorithm with $n\_estimators$ Parameter Variations for Liver Disease Classification*

Ersa Muliani<sup>1\*</sup>, Asmarani Ayudhia<sup>2</sup>, Maulidya<sup>3</sup>, Novan Alkaf Bahraini Saputra<sup>4</sup>,  
Nuruddin Wiranda<sup>5</sup>

<sup>1,2,3,4,5</sup>Prodi Pendidikan Komputer, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Lambung Mangkurat, Banjarmasin, Indonesia

Email: <sup>1</sup>[ersamuliani14@gmail.com](mailto:ersamuliani14@gmail.com), <sup>2</sup>[asmrniaydhia@gmail.com](mailto:asmrniaydhia@gmail.com), <sup>3</sup>[mlidya212@gmail.com](mailto:mlidya212@gmail.com),  
<sup>4</sup>[novan.saputra@ulm.ac.id](mailto:novan.saputra@ulm.ac.id), <sup>5</sup>[nuruddin.wd@ulm.ac.id](mailto:nuruddin.wd@ulm.ac.id)

\*Corresponding Author

#### ABSTRAK

Penyakit hati merupakan gangguan serius yang dapat menurunkan kualitas hidup dan menimbulkan komplikasi jika tidak terdeteksi sejak dini. Penelitian ini mengevaluasi penerapan algoritma *Random Forest* untuk klasifikasi penyakit hati menggunakan dataset *Indian Liver Patient Records*. Algoritma ini dipilih karena mampu menangani data kompleks dan mengurangi *overfitting*. Pengujian dilakukan dengan variasi parameter  $n\_estimators$  (100, 200, 300) serta teknik *10-fold cross-validation*. Evaluasi model mencakup akurasi, presisi, dan *recall*. Penelitian ini juga menerapkan teknik SMOTE untuk mengatasi ketidakseimbangan kelas dalam data, yang berkontribusi terhadap peningkatan performa model. Dibandingkan dengan studi sebelumnya yang menggunakan *Random Forest* tanpa tuning parameter, penelitian ini menunjukkan peningkatan akurasi sebesar 7%. Hasil terbaik diperoleh pada  $n\_estimators = 200$  dengan akurasi 80,17%, presisi 80,28%, dan *recall* 80,17%. Temuan ini menunjukkan bahwa algoritma *Random Forest* dengan parameter yang dioptimalkan efektif dalam mendeteksi penyakit hati dan dapat mendukung pengambilan keputusan medis secara lebih akurat.

**Kata Kunci:** *Klasifikasi; Machine Learning; Penyakit Hati; Random Forest*

#### ABSTRACT

*Liver disease is a serious condition that can reduce quality of life and cause complications if not detected early. This study evaluates the application of the Random Forest algorithm for liver disease classification using the Indian Liver Patient Records dataset. Random Forest was selected for its ability to handle complex data and reduce overfitting. The experiment tested*



*n\_estimators* values (100, 200, 300) with 10-fold cross-validation. Evaluation metrics included accuracy, precision, and recall. SMOTE was also applied to handle class imbalance, contributing to better model performance. Compared to previous studies that applied Random Forest without parameter tuning, this study achieved a 7% improvement in accuracy. The best performance was achieved at *n\_estimators* = 200, with 80.17% accuracy, 80.28% precision, and 80.17% recall. These results indicate that Random Forest with optimized parameters is effective in detecting liver disease and can support more accurate medical decision-making.

**Keywords:** Classification; Liver Disease; Machine Learning; Random Forest

---

## 1. PENDAHULUAN

Penyakit hati adalah masalah kesehatan serius di seluruh dunia. Jumlah penderitanya terus bertambah, dan hal ini sangat memengaruhi kualitas hidup mereka. Penyakit ini dapat berkembang menjadi kondisi yang lebih parah seperti sirosis atau kanker hati, bahkan dapat menyebabkan kematian jika tidak segera terdeteksi dan ditangani [1]. Karena risikonya yang dapat berkembang menjadi kanker hati, sangat penting untuk memiliki metode deteksi dini yang dapat diandalkan [2][3]. Mendeteksi penyakit hati secara cepat dan tepat sangat penting agar dokter dapat memberikan penanganan yang sesuai, merencanakan perawatan yang efektif, dan pada akhirnya meningkatkan harapan hidup pasien. Oleh karena itu, pencarian metode yang lebih efisien dan akurat untuk mendeteksi penyakit hati merupakan bidang penelitian yang sangat penting dalam dunia kedokteran dan teknologi informasi.

Dalam beberapa tahun terakhir, kemajuan pesat di bidang *machine learning* telah membuka peluang baru bagi diagnosis penyakit yang lebih canggih dan otomatis. Algoritma *machine learning* memiliki kemampuan untuk mempelajari pola kompleks dari data medis, yang sering kali sulit diidentifikasi secara manual oleh tenaga medis. Berbagai penelitian telah menunjukkan potensi *machine learning* dalam mendukung pengambilan keputusan klinis, termasuk dalam identifikasi penyakit hati [4] [3].

Salah satu algoritma yang menunjukkan kinerja menjanjikan dalam tugas klasifikasi adalah *Random Forest*. Algoritma ini merupakan metode *ensemble learning* yang membangun sejumlah pohon keputusan selama tahap pelatihan dan menghasilkan output berupa kelas yang paling sering muncul (untuk klasifikasi) atau rata-rata prediksi (untuk regresi) dari seluruh pohon [5]. Keunggulan *Random Forest* terletak pada kemampuannya dalam menangani dataset berdimensi tinggi, mengurangi *overfitting*, dan menghasilkan tingkat akurasi yang tinggi [1]. Penerapan *Random Forest* juga telah terbukti berhasil dalam konteks penyakit hati lainnya, seperti prediksi infeksi virus Hepatitis C [6] maupun prediksi sirosis hati [7], yang menunjukkan fleksibilitas dan keandalan algoritma ini.

Parameter seperti (jumlah pohon dalam hutan) dan teknik validasi silang seperti *K-Fold Cross-Validation* berperan penting dalam mengoptimalkan performa model [8]. Penyesuaian nilai *n\_estimators* dapat memengaruhi kompleksitas model serta kemampuannya dalam menangkap pola data, sementara *K-Fold Cross-Validation* memastikan bahwa model dievaluasi secara andal dan konsisten pada berbagai subset data [9][10]. Secara khusus, *K-Fold Cross-Validation* membagi dataset menjadi *k* lipatan yang sama besar, menggunakan *k-1* lipatan untuk pelatihan dan satu lipatan untuk pengujian, yang diulang sebanyak *k* kali guna memperoleh evaluasi yang lebih representatif terhadap kinerja model [9][11].

Dalam penelitian ini, pemilihan nilai  $n\_estimators$  sebanyak 100, 200, dan 300 didasarkan pada pertimbangan umum dalam praktik *machine learning* untuk algoritma *Random Forest*. Nilai 100 sering kali digunakan sebagai titik awal yang baik karena umumnya sudah cukup untuk mencapai performa yang stabil dan mengurangi varians. Sementara itu, nilai 200 dan 300 dipilih untuk mengeksplorasi apakah penambahan jumlah pohon dapat secara signifikan meningkatkan akurasi tanpa menyebabkan peningkatan beban komputasi yang berlebihan atau risiko *overfitting* yang tidak diinginkan [12]. Peningkatan  $n\_estimators$  memang cenderung meningkatkan akurasi hingga titik tertentu, setelah itu peningkatannya menjadi sangat kecil atau tidak signifikan. Oleh karena itu, rentang nilai ini dipilih untuk menemukan titik optimal yang seimbang antara performa dan efisiensi komputasi, serta untuk mengonfirmasi tren kinerja model.

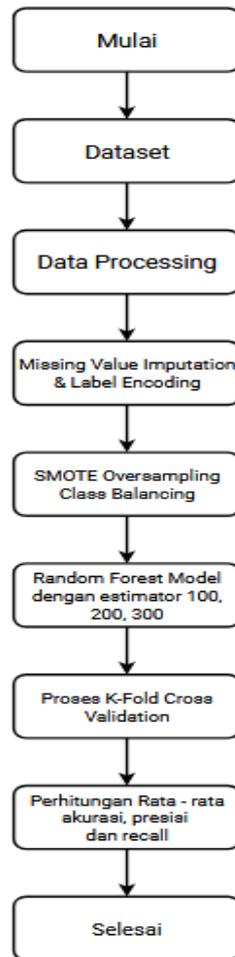
Penelitian ini akan menggunakan dataset *Indian Liver Patient Records* yang tersedia secara publik di UCI Machine Learning Repository [2]. Dataset ini merupakan sumber data yang umum digunakan dalam penelitian klasifikasi penyakit hati, sehingga memungkinkan perbandingan hasil dengan studi sebelumnya serta mendukung validasi model yang lebih kuat. Berdasarkan potensi tersebut, penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi algoritma *Random Forest* sebagai metode klasifikasi penyakit hati dengan menggunakan dataset *Indian Liver Patient Records*. Penelitian ini akan menganalisis pengaruh variasi parameter  $n\_estimators$  (100, 200, 300) dan teknik *K-Fold Cross-Validation* ( $k = 10$ ) terhadap kinerja model, dengan fokus pada metrik akurasi, presisi, dan *recall*. Hasil dari penelitian ini diharapkan dapat memberikan konfigurasi model *Random Forest* yang optimal untuk deteksi penyakit hati, sehingga dapat berkontribusi pada pengambilan keputusan medis yang lebih akurat dan terinformasi.

Tujuan penelitian ini adalah mengimplementasikan algoritma *Random Forest* sebagai metode klasifikasi penyakit hati, dengan fokus pada pengujian variasi parameter  $n\_estimators$  dan teknik *K-Fold Cross-Validation*. Studi ini merupakan pengembangan dari penelitian sebelumnya yang menunjukkan bahwa model *Random Forest* tanpa penyaringan fitur menghasilkan akurasi terbaik sebesar 73% dalam memprediksi penyakit hati, dibandingkan dengan model lain seperti Regresi Logistik, *Decision Tree*, *Gradient Boosting Decision Trees*, *AdaBoost*, *XGBoost*, dan *Deep Neural Network*. Tujuan utama penelitian ini adalah untuk membuktikan bahwa modifikasi algoritma *Random Forest* melalui pengujian nilai  $n\_estimators$  (100, 200, 300) dan *K-Fold Cross-Validation* ( $k = 10$ ) dapat meningkatkan akurasi model di atas 73% pada dataset yang digunakan. Dengan demikian, kontribusi spesifik dari penelitian ini menjadi jelas meskipun *Random Forest* telah diterapkan sebelumnya, tapi belum ada studi yang secara sistematis menguji variasi parameter  $n\_estimators$  pada dataset *Indian Liver Patient Records* (ILPD) dengan mengintegrasikan teknik penyeimbangan kelas SMOTE dan evaluasi menggunakan *K-Fold Cross-Validation* untuk mencapai optimasi performa seperti yang diusulkan dalam penelitian ini.

## 2. METODE PENELITIAN

Penelitian ini termasuk dalam kategori penelitian kuantitatif karena menggunakan data numerik pasien dengan penyakit hati yang dianalisis menggunakan algoritma *machine learning*, yaitu *Random Forest*. Data tersebut diproses melalui tahapan pra-pemrosesan, penyeimbangan data, serta pelatihan model untuk menghasilkan klasifikasi. Evaluasi model

dilakukan dengan menggunakan metrik akurasi, presisi, dan *recall*. Gambar 1 menyajikan tahapan proses dalam penelitian ini.



Gambar 1. Alur Metode yang Digunakan

## 2.1. Dataset

Data yang digunakan dalam penelitian ini adalah dataset *Indian Liver Patient Records* yang diperoleh dari Kaggle [2]. Dataset ini terdiri dari 582 data pasien (instance) dengan 10 atribut fitur klinis, seperti usia, jenis kelamin, total bilirubin, dan rasio albumin/globulin, serta satu atribut target yang mengklasifikasikan kondisi pasien [1].

Table 1. Informasi *Indian Liver Patient Records* Dataset

Atribut	Domain
Age	4-90 (umur pasien)
Gender	1 = male, 2=female
Total Bilirubin	0,4-75 (kadar bilirubin total)
Direct Bilirubin	0,1-19,7 (kadar bilirubin langsung)
Alkaline Phosphotase	63-2110 (kadar fosfatase alkali)
Alamine Amino transferase	10-2000 (kadar aminotransferase alamin)
Aspartate Amino transferase	10-4929 (kadar aminotransferase alamin)
Total Protiens	2,7 - 9,6 (kadar total protein)

Atribut	Domain
Albumin	0,9 - 5,5 (kadar albumin)
Albumin and Globulin Ratio	0,3 - 2,8 (rasio albumin dan globulin)
Dataset	1 = penyakit hati, 2 = tidak penyakit hati

## 2.2. Pra-Pemrosesan Data

Sebelum data digunakan untuk pelatihan model, dilakukan tahap pra-pemrosesan. Tahap ini bertujuan untuk membersihkan data dari nilai yang hilang (*missing values*) maupun data yang berlebihan (*redundant values*) [3][8]. Nilai yang hilang pada atribut tertentu ditangani dengan menggantinya menggunakan nilai terdekat atau rata-rata dari fitur tersebut guna menjaga integritas dataset [3][13].

## 2.3. Pembagian Dataset

Setelah data dibersihkan, evaluasi performa model tidak dilakukan dengan pembagian data latih dan uji secara statis, melainkan dengan menerapkan metode *K-Fold Cross-Validation* untuk memperoleh hasil evaluasi yang lebih stabil dan andal [9]. Dalam teknik ini, dataset dibagi menjadi  $K$  bagian (*fold*) yang sama besar. Model akan dilatih dan diuji sebanyak  $K$  kali, di mana pada setiap iterasi satu *fold* digunakan sebagai data uji dan  $K-1$  *fold* sisanya digunakan sebagai data latih [10]. Kinerja akhir model kemudian diukur berdasarkan rata-rata dari hasil  $K$  iterasi tersebut [11].

## 2.4. Implementasi Model

Berbeda dengan penelitian sebelumnya, penelitian ini hanya berfokus pada implementasi Algoritma Random Forest. Random Forest adalah metode *ensemble learning* yang bekerja dengan membangun banyak pohon keputusan (*decision tree*) dan menggabungkan hasilnya untuk mendapatkan prediksi yang lebih akurat dan stabil [3][5][14]. Sebagai metode *ensemble*, Random Forest secara inheren dirancang untuk mengatasi beberapa kelemahan yang ditemukan pada pohon keputusan tunggal, seperti kecenderungan untuk *overfitting* [16].

Untuk memperoleh performa model yang optimal, penelitian ini melakukan pengujian dengan memvariasikan *hyperparameter* utama dari *Random Forest*, yaitu  $n\_estimators$  [4]. Proses penyesuaian *hyperparameter* ini sangat penting karena dapat memengaruhi kinerja prediksi model secara signifikan [12]. Parameter  $n\_estimators$  merujuk pada jumlah pohon keputusan yang akan dibangun dalam model untuk proses prediksi [4]. Selain itu, untuk memastikan evaluasi model yang lebih kuat dan menghindari bias akibat pembagian data tunggal, penelitian ini menerapkan teknik *k-fold cross-validation* [1][4][13]. Dalam metode ini, data latih dibagi menjadi  $k$  bagian (*folds*) yang sama besar. Model kemudian dilatih sebanyak  $k$  kali, di mana pada setiap iterasi satu bagian digunakan sebagai data validasi, sementara bagian lainnya digunakan sebagai data latih [1][13]. Hasil performa dari setiap iterasi kemudian dirata-ratakan untuk memperoleh evaluasi akhir yang lebih andal [1][13].

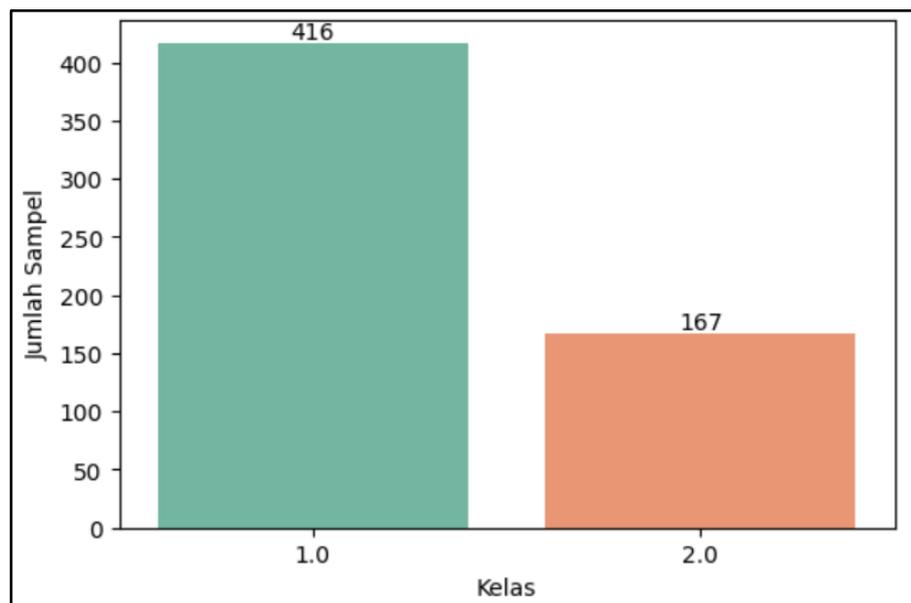
Penelitian ini juga mengatasi masalah ketidakseimbangan kelas pada dataset dengan menggunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE). SMOTE bekerja dengan menghasilkan sampel sintesis baru dari kelas minoritas untuk menyeimbangkan distribusi data, sehingga dapat meningkatkan kinerja metode klasifikasi. Teknik ini sangat relevan karena dataset medis seperti *Indian Liver Patient Records* (UCI Machine Learning Repository, "Liver Disorder Data Set"), yang sering digunakan dalam klasifikasi penyakit hati,

umumnya memiliki distribusi data yang tidak seimbang antara pasien yang menderita penyakit hati dan yang tidak [10][15].

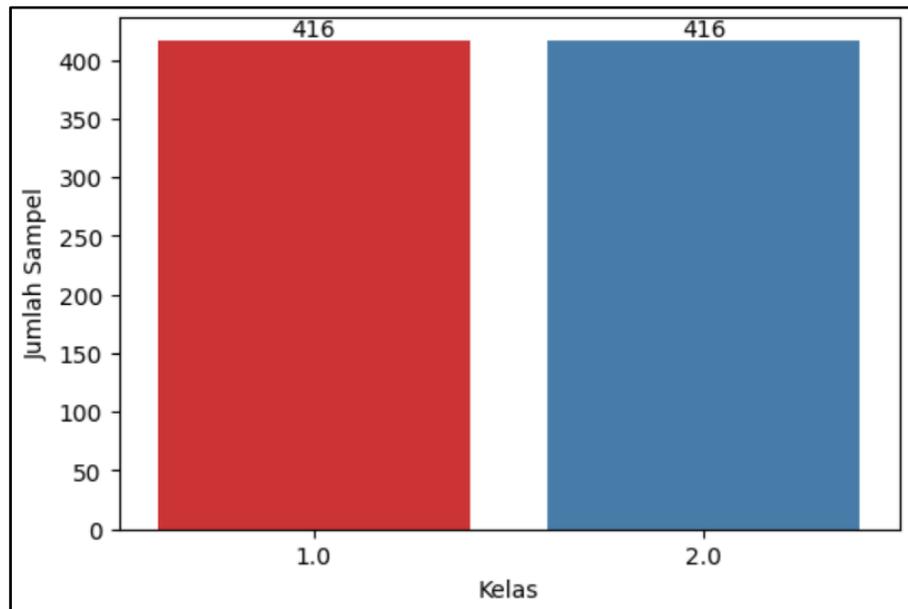
Dalam penelitian ini, metode evaluasi model yang digunakan adalah *10-fold cross-validation*. Pemilihan nilai  $K = 10$  didasarkan pada posisinya sebagai standar umum yang telah terbukti efektif dan banyak diimplementasikan dalam berbagai penelitian *machine learning* [1][11]. Penggunaan konfigurasi ini memberikan beberapa keuntungan utama. Pertama, metode ini efektif untuk menjaga agar model tidak terlalu menyederhanakan permasalahan (*underfitting*) dan juga tidak sekedar menghafal data (*overfitting*), karena pada setiap iterasi sembilan bagian data (*fold*) digunakan untuk melatih model, dan satu bagian sisanya digunakan untuk validasi. Dengan demikian, evaluasi yang dihasilkan tidak bersifat kebetulan dan lebih dapat diandalkan [9]. Selain itu, karena metode ini merupakan standar yang umum digunakan, hasil penelitian ini dapat dibandingkan secara adil dengan hasil dari penelitian lain yang mengangkat topik serupa [7].

### 3. HASIL DAN PEMBAHASAN

Ketidakseimbangan kelas pada dataset *Indian Liver Patient Records* menjadi tantangan utama dalam membangun model klasifikasi yang akurat. Sebelum dilakukan *oversampling*, jumlah data pasien dengan penyakit hati (kelas 1) berjumlah 137, sedangkan pasien tanpa penyakit hati (kelas 2) sebanyak 329. Ketidakseimbangan ini dapat menyebabkan model lebih cenderung memprediksi kelas mayoritas, sehingga menurunkan sensitivitas terhadap kelas minoritas. Untuk mengatasi hal tersebut, penelitian ini menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE) yang terbukti mampu menyeimbangkan distribusi data dan meningkatkan kinerja klasifikasi [10].



Gambar 2. Sebelum Dilakukan Oversampling



Gambar 3. Sesudah Dilakukan Oversampling

Visualisasi pada Gambar 2 menunjukkan dominasi kelas mayoritas sebelum penerapan SMOTE. Sementara itu, Gambar 3 memperlihatkan hasil setelah SMOTE diterapkan, di mana distribusi data menjadi lebih seimbang antara kedua kelas. Distribusi ini penting untuk mengurangi bias klasifikasi dan meningkatkan kemampuan model dalam mengenali pola pada kedua kelas [13].

Setelah data di seimbangkan, model Random Forest diuji dengan variasi parameter  $n\_estimators$  (100, 200, dan 300) menggunakan teknik 10-fold cross-validation. Teknik ini digunakan untuk menghindari hasil evaluasi yang bias akibat pembagian data tunggal dan telah banyak diterapkan dalam penelitian klasifikasi berbasis machine learning [9].

Tabel 2. Evaluasi Performa Model *Random Forest* dengan  $n\_estimators = 100, 200,$  dan 300 menggunakan 10-fold cross-validation

$n\_estimators$	Akurasi	Presisi	Recall	lata-rata
100	79,57	79,68	79,57	79,61
200	80,17	80,28	80,17	80,21
300	79,93	80,22	79,93	80,03

Hasil evaluasi menunjukkan bahwa pada  $n\_estimators = 100$ , model menghasilkan akurasi sebesar 79,57%, presisi 79,68%, dan *recall* 79,57%. Saat jumlah pohon ditingkatkan menjadi 200, terjadi peningkatan akurasi menjadi 80,17%, dengan presisi 80,28% dan *recall* 80,17%. Namun, ketika jumlah pohon dinaikkan lebih lanjut menjadi 300, performa model cenderung stagnan pada akurasi 79,93%, dengan presisi 80,22% dan *recall* 79,93%.

Peningkatan performa model dari  $n\_estimators = 100$  ke 200 menunjukkan bahwa penambahan jumlah pohon mampu memperkuat kemampuan prediktif model melalui reduksi varians prediksi [5]. Algoritma ensemble seperti Random Forest menunjukkan kinerja yang lebih baik dibandingkan model individual dalam prediksi penyakit hati [7]. Namun, ketika jumlah  $n\_estimators$  melebihi titik tertentu, peningkatan performa menjadi tidak signifikan, sebagaimana diamati pada  $n\_estimators = 300$ . Fenomena ini menunjukkan adanya titik

optimal pada parameter tersebut, di mana penambahan lebih lanjut hanya meningkatkan kompleksitas model tanpa memberikan dampak nyata terhadap akurasi.

Pengaturan  $n\_estimators$  yang tepat sangat penting untuk menghindari redundansi informasi dan pemborosan komputasi [12]. Dalam penelitian ini, nilai optimal ditemukan pada  $n\_estimators = 200$ , yang memberikan keseimbangan terbaik antara akurasi dan efisiensi.

Selain itu, hasil penelitian ini menunjukkan adanya peningkatan akurasi sebesar 7% dibandingkan dengan penelitian sebelumnya, yang melaporkan akurasi sebesar 73% pada model Random Forest tanpa tuning parameter dan tanpa penyeimbangan kelas [4]. Hal ini menegaskan bahwa kombinasi antara teknik SMOTE, pengaturan parameter, dan validasi silang dapat secara signifikan meningkatkan performa klasifikasi penyakit hati.

#### 4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan dan mengevaluasi algoritma Random Forest untuk klasifikasi penyakit hati menggunakan dataset Indian Liver Patient Records, secara langsung menjawab permasalahan mengenai kebutuhan deteksi penyakit hati yang lebih akurat. Kami berfokus pada optimasi performa model dengan menguji berbagai konfigurasi  $n\_estimators$  dan *K-Fold Cross-Validation*. Hasil evaluasi menunjukkan bahwa konfigurasi  $n\_estimators$  dan *K-Fold Cross-Validation* memiliki dampak positif signifikan terhadap akurasi klasifikasi. Secara spesifik, model Random Forest mencapai kinerja terbaik dengan  $n\_estimators = 200$ , menghasilkan akurasi sebesar 80,17%, presisi 80,28%, dan *recall* 80,17%. Angka ini merupakan peningkatan akurasi sebesar 7% dibandingkan dengan akurasi 73% yang menjadi target awal, sebagaimana disebutkan dalam penelitian sebelumnya mengenai prediksi penyakit hati. Konsistensi hasil evaluasi melalui *K-Fold Cross-Validation* menegaskan keandalan model dalam mendeteksi penyakit hati, yang merupakan solusi krusial untuk mendukung pengambilan keputusan medis yang lebih akurat dan terinformasi. Selain itu, penanganan ketidakseimbangan kelas menggunakan metode *oversampling* juga terbukti berkontribusi pada peningkatan performa model secara keseluruhan, memastikan model dapat belajar secara efektif dari kedua kelas data.

Meskipun hasil ini sangat menjanjikan, penelitian selanjutnya dapat berfokus pada validasi dan optimasi model lebih lanjut. Sangat direkomendasikan untuk menguji model ini pada dataset yang lebih besar dan lebih beragam, yang mencakup data dari berbagai populasi atau wilayah geografis. Hal ini penting untuk memastikan sistem model bisa diandalkan saat diterapkan di dunia nyata dan untuk menjamin performanya agar tidak bias.

#### DAFTAR PUSTAKA

- [1] I. Herliawan *et al.*, "Classification of Liver Disease by Applying Random Forest Algorithm and Backward Elimination," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 6, no. 1, pp. 89–94, 2020, doi: <https://doi.org/10.33480/jitk.v6i1.1424>
- [2] UCI Machine Learning Repository, "Liver disorder data set," <https://archive.ics.uci.edu/ml/datasets/Liver+Disorders>.
- [3] K. Biju, M. B. Manoj, and Mohan. H, "Diagnosis of Chronic Liver Disease using Machine Learning Techniques," *Int J Res Appl Sci Eng Technol*, vol. 11, no. 6, pp. 346–351, Jun. 2023, doi: <https://doi.org/10.17762/itii.v9i2.382>.

- [4] H. Wei, "A Random Forest-based Prediction for Liver Cirrhosis," *Applied and Computational Engineering*, vol. 67, no. 1, pp. 98–106, Jul. 2024, doi: <https://doi.org/10.54254/2755-2721/78/20240642>.
- [5] K. Widya Kayohana, "Klasifikasi Penyakit Hati Menggunakan Random Forest dan KNN," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 4, pp. 7924–7929, 2024, doi: <https://doi.org/10.36040/jati.v8i4.10457>.
- [6] A. Rozy, "Penerapan Random Forest Untuk Prediksi Virus Hepatitis C," *FIMERKOM: Journal of Information Systems and Technology*, vol. 1, no. 1, pp. 19–23, 2024, doi: <https://doi.org/10.36040/jati.v8i4.10457>.
- [7] M. Ghosh *et al.*, "A comparative analysis of machine learning algorithms to predict liver disease," *Intelligent Automation and Soft Computing*, vol. 30, no. 3, pp. 917–928, 2021, doi: <https://doi.org/10.32604/iasc.2021.017989>.
- [8] W. Andriyani *et al.*, "Analisis Data Penerimaan Peserta Didik Baru Menggunakan Cross Validation Dan Algoritma Decision Tree Di SMA Negeri 1 Bandung," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 3, 2024, doi: <https://doi.org/10.36040/jati.v8i3.9603>.
- [9] W. Wijiyanto, A. I. Pradana, S. Sopingi, and V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," *Jurnal Algoritma*, vol. 21, no. 1, May 2024, doi: <https://doi.org/10.33364/algoritma/v.21-1.1618>
- [10] Y. Khoeri, R. Kurniawan, and Y. A. Wijaya, "Penerapan Teknik SMOTE Dan Cross Validation Pada Decision Tree Untuk Klasifikasi Tingkat Kemacetan Lalu Lintas," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 2, 2024, doi: <https://doi.org/10.36040/jati.v8i2.8879>.
- [11] A. P. Windarto, S. Defit, and A. Wanto, "Optimalisasi Parameter dengan Cross Validation dan Neural Back-propagation Pada Model Prediksi Pertumbuhan Industri Mikro dan Kecil," *JURNAL SISTEM INFORMASI BISNIS*, vol. 11, no. 1, pp. 34–42, Jun. 2021, doi: <https://doi.org/10.21456/vol11iss1pp34-42>.
- [12] P. Contreras, J. Orellana-Alvear, P. Muñoz, J. Bendix, and R. Célleri, "Influence of random forest hyperparameterization on short-term runoff forecasting in an andean mountain catchment," *Atmosphere (Basel)*, vol. 12, no. 2, Feb. 2021, doi: <https://doi.org/10.3390/atmos12020238>
- [13] Y. Cahyono, P. Rosyani, F. S. Syah, F. S. Putri, I. Ashari, and K. Sofian, "Liver Disease Classification Using Decision Tree and Random Forest Algorithms," *International Journal of Integrative Sciences*, vol. 4, no. 1, pp. 135–140, Feb. 2025, doi: <https://doi.org/10.55927/ijis.v4i1.13509>.
- [14] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Stat Surv*, vol. 4, pp. 40–79, 2010, doi: <https://doi.org/10.1214/09-SS054>.
- [15] R. Meidiansyah and M. Herdayati, "Risk of Liver Cancer in Patients with Liver Disease (BPJS Sample Data Analysis)," *Jurnal Kesehatan Masyarakat Mulawarman*, vol. 4, p. 135, 2024, doi: <https://doi.org/10.55927/ijis.v4i1.13509>.