

**ANALISIS PERFORMA ALGORITMA RANDOM FOREST  
DALAM MEMPREDIKSI PENYAKIT LIVER  
MENGUNAKAN DATASET DARI KAGGLE**

**SKRIPSI**

**Diajukan sebagai Salah Satu Syarat untuk Memperoleh Kelulusan  
Jenjang Strata Satu pada Program Studi Teknik Informatika**

**Oleh**

**Marya Ovi Sesa**

**2255201152**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH BENGKULU  
2026**

**HALAMAN PERSETUJUAN**

**ANALISIS PERFORMA ALGORITMA RANDOM FOREST  
DALAM MEMPREDIKSI PENYAKIT LIVER  
MENGUNAKAN DATASET DARI KAGGLE**

Oleh  
Marya Ovi Sesa  
2255201152

Tugas Akhir ini telah Diterima dan Disahkan  
untuk Memenuhi Persyaratan Mencapai Gelar  
SARJANA KOMPUTER (S.Kom)

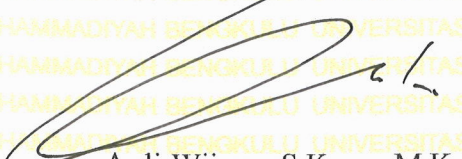
Pada  
PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK UNIVERSITAS MUHAMMADIYAH BENGKULU

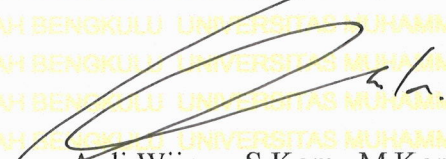
Bengkulu, 11 Februari 2026

Disetujui oleh

Ketua Program Studi,

Dosen Pembimbing,

  
Ardi Wijaya, S.Kom., M.Kom  
NP. 19880511 201408 1 181

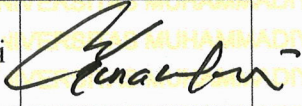

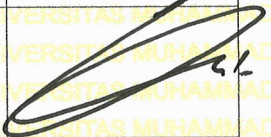
  
Ardi Wijaya, S.Kom., M.Kom  
NIDN.0211058803

**LEMBAR PERSETUJUAN HASIL REVISI**  
**ANALISIS PERFORMA ALGORITMA RANDOM FOREST**  
**DALAM MEMPREDIKSI PENYAKIT LIVER**  
**MENGGUNAKAN DATASET DARI KAGGLE**

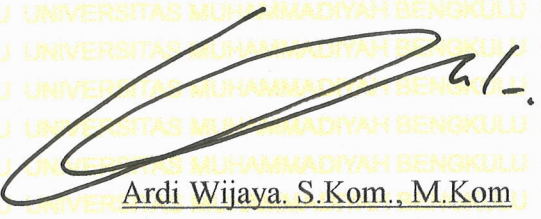
Oleh  
Marya Ovi Sesa  
2255201152

Telah Melakukan Revisi Sesuai dengan Perubahan  
dan Perbaikan yang Diminta Pada Saat Sidang Tugas Akhir.

Bengkulu, 08 Maret 2026  
Menyetujui

No	Nama Dosen	Keterangan	Tanda Tangan
1	Gunawan, S.Kom., M.Kom.	Ketua Penguji	
2	Dedy Abdullah, S.T., M.Eng.	Penguji 1	
3	Ardi Wijaya, S.Kom., M.Kom.	Penguji 2	

Mengetahui  
Ketua Program Studi Teknik Informatika

  
Ardi Wijaya, S.Kom., M.Kom  
NP. 19880511 201408 1 181

## LEMBAR PENGESAHAN

# ANALISIS PERFORMA ALGORITMA RANDOM FOREST DALAM MEMPREDIKSI PENYAKIT LIVER MENGUNAKAN DATASET DARI KAGGLE

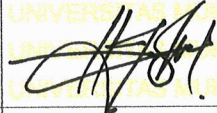
### SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh Kelulusan  
Jenjang Strata Satu pada Program Studi Teknik Informatika

Oleh

Marya ovi sesa  
2255201152

Bengkulu, 08 Maret 2026

No	Nama Dosen	Keterangan	Tanda Tangan
1	Gunawan, S.Kom., M.Kom.	Ketua Penguji	
2	Dedy Abdullah, S.T., M.Eng.	Penguji 1	
3	Ardi Wijaya, S.Kom., M.Kom.	Penguji 2	

Mengesahkan  
Dekan Fakultas Teknik

  
RG Guntur Alam, M.Kom., Ph.D  
NP. 19730101 200004 1 040

## SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa:

- 1) Naskah Skripsi ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik, baik di Universitas Muhammadiyah Bengkulu maupun perguruan tinggi lainnya.
- 2) Skripsi ini murni merupakan karya penelitian saya sendiri dan tidak menjiplak karya pihak lain. Dalam hal ada bantuan atau arahan dari pihak lain maka telah saya sebutkan identitas dan jenis bantuannya di dalam lembar ucapan terima kasih.
- 3) Seandainya ada karya pihak lain yang ternyata memiliki kemiripan dengan karya saya ini, maka hal ini adalah di luar pengetahuan saya dan terjadi tanpa kesengajaan dari pihak saya.

Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terbukti adanya kebohongan dalam pernyataan ini, maka saya bersedia menerima sanksi akademik sesuai norma yang berlaku di Universitas Muhammadiyah Bengkulu.

Bengkulu, 08 Maret 2026



Marya Ovi Sesa  
2255201152

## **DAFTAR RIWAYAT HIDUP**

### **Identitas Pribadi**

Nama : Marya Ovi Sesa  
TTL : Gunung Tiga, 23 Maret 2004  
Agama : Islam  
Anak Ke : 2  
Alamat : Desa Gunung Tiga 1 Kecamatan, Semidang  
Gumay, Kab.Kaur, Bengkulu  
Email : maryaovisesa93@gmail.com

### **Nama Orang Tua**

Ayah : Ahmad Mansyur  
Pekerjaan : Buruh Bangunan  
Ibu : Eli Hasnawati  
Pekerjaan : Ibu Rumah Tangga

### **Riwayat Pendidikan**

1. SDN 63 KAUR : 2011-2016
2. MTsN 2 KAUR : 2016-2018
3. SMA N 3 KAUR : 2019-2022
4. Universitas Muhammadiyah Bengkulu : 2022-2026

## **MOTO**

Allah Memang Tidak Menjanjikan Hidupmu Akan Selalu Mudah, Tapi Dua Kali

Allah Berjanji Bahwa : “Fa Inna Ma’al – ‘Usri Yusra, Inna Ma’al – ‘Usri Yusra”

“Setiap Kesulitan Pasti Ada Kemudahan”

**(Q.S. Al-Insyirah 94:5-6)**

“aku membahayakan nyawa ibuku untuk lahir ke dunia, jadi tidak mungkin aku tidak ada artinya, dan aku membuat ayahku bekerja setiap hari hingga lelah, jadi aku pastikan lelahnya tidak sia-sia”

**(Penulis)**

## ABSTRAK

# ANALISIS PERFORMA ALGORITMA RANDOM FOREST DALAM MEMPREDIKSI PENYAKIT LIVER MENGUNAKAN DATASET DARI KAGGLE

Nama : Marya ovi sesa  
NPM : 2255201152  
Dosen Pembimbing : Ardi Wijaya. S.Kom., M.Kom

Penyakit liver merupakan gangguan kesehatan yang dapat menurunkan fungsi metabolisme tubuh dan berpotensi menyebabkan komplikasi serius apabila tidak terdeteksi sejak dini. Penelitian ini bertujuan untuk menganalisis efektivitas algoritma Random Forest dalam memprediksi penyakit liver berdasarkan data medis pasien. Dataset yang digunakan berasal dari platform Kaggle dan terdiri dari 580 data pasien dengan 11 variabel yang mencakup data demografis dan hasil pemeriksaan laboratorium. Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimen yang meliputi tahapan pengumpulan data, preprocessing data berupa penghapusan data duplikasi dan encoding data kategorikal, serta pembagian data latih dan data uji dengan rasio 80:20. Model prediksi dibangun menggunakan algoritma Random Forest dengan 100 pohon keputusan dan dilatih menggunakan data latih. Evaluasi performa model dilakukan menggunakan confusion matrix, accuracy, precision, recall, F1-score, dan Area Under Curve (AUC). Hasil penelitian menunjukkan bahwa algoritma Random Forest menghasilkan nilai akurasi sebesar 74% dan nilai AUC sebesar 0,797, yang menunjukkan kemampuan klasifikasi yang cukup baik. Model memiliki kemampuan yang sangat baik dalam mendeteksi pasien dengan penyakit liver, ditunjukkan oleh nilai recall kelas positif sebesar 0,93 dan F1-score sebesar 0,84, meskipun performa pada kelas non-liver masih terbatas. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma Random Forest cukup efektif dan andal dalam memprediksi penyakit liver berbasis data medis serta berpotensi digunakan sebagai sistem pendukung keputusan medis.

**Kata kunci:** Penyakit Liver, Random Forest, *Machine Learning*.

**ABSTRACT**  
**ANALISIS PERFORMA ALGORITMA RANDOM FOREST DALAM**  
**MEMPREDIKSI PENYAKIT LIVER MENGGUNAKAN DATASET DARI**  
**KAGGLE**

Name : Marya Ovi Sesa  
Student ID : 2255201152  
Supervisor : Ardi Wijaya. S.Kom., M.Kom

Liver disease is a health disorder that can reduce the body's metabolic function and potentially lead to serious complications if not detected early. This study aims to analyze the effectiveness of the Random Forest algorithm in predicting liver disease based on patients' medical data. The dataset used in this study was obtained from the Kaggle platform and consists of 580 patient records with 11 variables, including demographic data and laboratory test results. This research employed a quantitative method with an experimental approach consisting of several stages, namely data collection, data preprocessing including duplicate data removal and categorical data encoding, and data splitting into training and testing sets with an 80:20 ratio. The prediction model was developed using the Random Forest algorithm with 100 decision trees and trained using the training dataset. The model performance was evaluated using a confusion matrix, accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). The results show that the Random Forest algorithm achieved an accuracy of 74% and an AUC value of 0.797, indicating a fairly good classification performance. The model demonstrates a very strong ability to detect patients with liver disease, as indicated by a positive class recall value of 0.93 and an F1-score of 0.84, although the performance in identifying the non-liver class is still limited. Based on these results, it can be concluded that the Random Forest algorithm is sufficiently effective and reliable in predicting liver disease based on medical data and has the potential to be used as a medical decision support system.

**Keywords:** *Liver Disease, Random Forest, Machine Learning.*

## KATA PENGANTAR

Assalamu'alaikum, wr. wb

Puji syukur penulis panjatkan kehadiran Allah Subhanahu Wa Ta'ala, atas berkat, rahmat, ridho dan karunia-Nya sehingga penulis dapat menyelesaikan proposal skripsi yang berjudul **“ANALISIS ALGORITMA RANDOM FOREST DALAM MEMREDIKSI PENYAKIT LIVER MENGGUNAKAN DATASET DARI KAGGLE”**. Shalawat dan salam senantiasa tercurahkan kepada junjungan kita Nabi Muhammad Shallallahu Alaihi Wasallam yang selalu menjadi tauladan bagi kita semua. Penulis mengucapkan terima kasih kepada semua pihak atas bantuan dan bimbingan dalam pembuatan skripsi ini, pada kesempatan ini penulis mengucapkan rasa terimakasih yang sebesar-besarnya kepada:

1. Bapak Dr. Susiyanto, MSi selaku Rektor Universitas Muhammadiyah Bengkulu
2. Bapak RG Guntur Alam, M.Kom, Ph.D selaku Dekan Fakultas Teknik Universitas Muhammadiyah Bengkulu
3. Bapak Ardi Wijaya, S.Kom., M.Kom selaku Kepala Program Studi Teknik Informatika Universitas Muhammadiyah Bengkulu
4. Bapak Ardi Wijaya, S.Kom., M.Kom selaku Dosen Pembimbing dalam penyusunan proposal skripsi.
5. Keluarga tercinta yang telah memberikan dukungan moral maupun materi.
6. Serta seluruh Dosen Fakultas Teknik Universitas Muhammadiyah Bengkulu.

Semoga Allah Subhanahu Wa Ta'ala selalu membalas apa yang mereka berikan untuk penulis dengan imbalan yang berlipat ganda. Penulis menyadari di dalam penulisan skripsi ini masih jauh dari kesempurnaan, untuk itu penulis mohon kritik, saran dan masukan yang bersifat membangun demi kesempurnaan penulisan skripsi ini. Sehingga dapat bermanfaat dan berguna untuk pengembangan ilmu pengetahuan, amin.

Wassalamu'alaikum, wr. wb

Bengkulu, Maret 2026

Marya ovi sesa  
NPM : 2255201152

## UCAPAN TERIMA KASIH

Puji syukur penulis ucapkan atas kehadiran Allah SWT karena berkat Rahmat dan karunia-Nya, penulis masih diberikan kesempatan untuk dapat menyelesaikan skripsi dengan judul “**analisis algoritma random forest dalam memprediksi penyakit liver menggunakan dataset dari kaggle**”, sebagai salah satu syarat untuk memperoleh gelar sarjana pada Fakultas Teknik Informatika, Universitas Muhammadiyah Bengkulu.

1. Kedua orang tuaku tersayang, Bapak Ahmad Mansyur dan Ibu Eli Hanawati. Terimakasih semua pengorbanan, peluh, usaha kasih, waktu doa, kesabaran dan ridho yang diberikan kepadaku
2. Bapak Dekan RG Guntur Alam, M.Kom., Ph.D selaku Dekan Fakultas Keguruan dan Ilmu Pendidikan Universitas Bengkulu.
3. Bapak Ardi Wijaya, S.Kom., M.Kom selaku Ketua Prodi Teknik Informatika Universitas Muhammadiyah Bengkulu.
4. Bapak Dandi sunardi, S.Kom., M.Kom selaku Dosen Pembimbing Akademik yang telah memberikan arahan dan bimbingan selama tiap semester
5. Bapak Ardi Wijaya, S.Kom., M.Kom selaku Dosen Pembimbing Utama skripsi saya yang telah membimbing saya dengan sabar dan teliti selama proses penulisan skripsi ini. Terima kasih pula atas segala ilmu, bantuan serta didikannya.

Bengkulu, 08 Maret 2026

Marya Ovi Sesa  
NPM : 2255201152

## DAFTAR ISI

Halaman

<b>HALAMAN JUDUL</b>	
<b>HALAMAN PERSETUJUAN</b> .....	<b>ii</b>
<b>HALAMAN PERSETUJUAN HASIL REVISI</b> .....	<b>iii</b>
<b>HALAMAN PENGESAHAN</b> .....	<b>iv</b>
<b>HALAMAN PERNYATAAN</b> .....	<b>v</b>
<b>HALAMAN DAFTAR RIWAYAT HIDUP</b> .....	<b>vi</b>
<b>MOTTO</b> .....	<b>vii</b>
<b>ABSTRAK</b> .....	<b>viii</b>
<b>ABSTRACT</b> .....	<b>ix</b>
<b>KATA PENGANTAR</b> .....	<b>xi</b>
<b>UCAPAN TERIMA KASIH</b> .....	<b>xii</b>
<b>DAFTAR ISI</b> .....	<b>xiii</b>
<b>DAFTAR GAMBAR</b> .....	<b>xv</b>
<b>DAFTAR TABEL</b> .....	<b>xvi</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang.....	1
1.1 Pertanyaan Penelitian.....	3
1.2 Tujuan Penelitian .....	3
1.4 Kerangka Kerja Penelitian .....	4
<b>BAB II TINJAUAN LITERATUR</b> .....	<b>6</b>
2.1 Penelitian Terdahulu.....	6
2.2 Penyakit Hati .....	8
2.3 Mesin Learning .....	8
2.4 Random Forest.....	8
2.5 Google Colab .....	8
<b>BAB III METODE PENELITIAN</b> .....	<b>9</b>
3.1 Analisis Masalah.....	9

3.2 Metode Penelitian .....	9
3.3 Pengumpulan Data.....	11
3.4 Pre-Processing .....	11
3.5 Data Preparations .....	12
3.6 Modeling.....	12
3.7 Evaluations .....	14
3.8 Tools Dan Perangkat Penelitian .....	14
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>	<b>27</b>
4.1 Hasil.....	27
4.2 Pembahasan .....	38
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>48</b>
5.1 Kesimpulan.....	48
5.2 Saran .....	48
<b>DAFTAR PUSTAKA</b>	

## DAFTAR GAMBAR

	Halaman
Gambar 4.1 Hasil Confusion Matrik .....	9
Gambar 4.2 Kurva Roc Model Random Forest .....	14
Gambar 4.3 Grafik Feature Prediksi Penyakit Liver .....	14

## DAFTAR TABEL

	Halaman
Tabel 1.1 <i>Framework</i> Penelitian.....	13
Tabel 1.2 <i>Literature Review</i> (SLR). ....	15
Tabel 3.1 Daftar Variable Dataset Penyakit Hati .....	15
Tabel 4.1 Struktur Dataset Penyakit Liver .....	16
Tabel 4.2 Hasil Pemeriksaan dan Penghapusan Data Duplikasi .....	16
Tabel 4.3 Hasil Missing <i>Value</i> .....	16
Tabel 4.4 Transformasi Label Variabel .....	17
Tabel 4.5 Pembagian Data Latih Dan Data Uji .....	13
Tabel 4.6 Modeling Random Forest. ....	15
Tabel 4.7 Hasil Evaluasi Model Random Forest .....	15
Tabel 4.8 Nilai Confusion Matrix Model Random Forest.....	16
Tabel 4.9 Hasil <i>Accuracy</i> .....	16
Tabel 4.10 Precision, Recal, Dan F1-Score .....	16
Tabel 4.11 Feature Importance Algoritma Random Forest .....	17

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Penyakit liver merupakan salah satu masalah kesehatan serius yang berdampak besar terhadap kualitas hidup masyarakat di seluruh dunia. Organ hati memiliki peran penting dalam proses metabolisme tubuh seperti detoksifikasi, penyimpanan energi, serta sintesis protein esensial. Kerusakan fungsi hati dapat menimbulkan gangguan sistemik dan berujung pada kondisi fatal seperti sirosis atau kanker hati. Oleh karena itu, deteksi dini terhadap penyakit liver menjadi hal yang sangat penting agar pengobatan dapat dilakukan sedini mungkin. Perkembangan teknologi informasi dan kecerdasan buatan (Artificial Intelligence/AI), khususnya di bidang machine learning, telah membuka peluang besar dalam mendukung proses diagnosis penyakit melalui analisis data medis secara lebih cepat dan akurat. (Chan et al., 2020)

Penelitian sebelumnya yang dilakukan oleh (Kesuma et al., 2023) pada artikel “Prediksi Penyakit Liver Menggunakan Algoritma Random Forest” menunjukkan bahwa model yang dibangun masih memiliki keterbatasan, terutama dalam hal variasi data dan generalisasi hasil. Penelitian tersebut menggunakan dataset lokal yang relatif kecil dan terbatas pada satu sumber, sehingga performa model belum dapat mencerminkan kemampuan prediksi yang stabil di berbagai kondisi populasi. Selain itu, penelitian tersebut belum mengeksplorasi potensi dataset yang lebih luas dan representatif, serta belum dilakukan validasi model yang mendalam untuk memastikan kehandalan hasil. Kesenjangan ini menjadi research gap yang penting untuk diteliti kembali.

Berdasarkan permasalahan tersebut, penelitian ini berfokus untuk menganalisis performa algoritma Random Forest dalam memprediksi penyakit liver menggunakan dataset dari Kaggle. Dataset ini dipilih karena memiliki jumlah data yang lebih besar, bersumber dari populasi yang beragam, serta telah banyak digunakan dalam penelitian internasional. Dalam implementasinya, penelitian ini memanfaatkan bahasa pemrograman Python sebagai tools utama,

karena memiliki pustaka analisis data dan machine learning yang lengkap seperti Scikit-learn, Pandas, dan Matplotlib. Melalui Python, seluruh proses mulai dari pra-pemrosesan data, pelatihan model, hingga evaluasi performa dapat dilakukan secara sistematis dan terukur menggunakan metrik seperti accuracy, precision, recall, F1-score, dan ROC-AUC. (Pedregosa et al., 2011)

Penelitian ini bertujuan untuk memberikan pemahaman lebih mendalam mengenai efektivitas algoritma Random Forest dalam mendeteksi penyakit liver berbasis data medis. Selain itu, hasil penelitian diharapkan dapat menjadi rujukan bagi pengembangan sistem pendukung keputusan medis berbasis machine learning yang mampu membantu tenaga kesehatan dalam proses diagnosis awal penyakit hati. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi ilmiah dalam bidang analisis data kesehatan, tetapi juga berpotensi memberikan manfaat praktis dalam peningkatan layanan medis berbasis teknologi.

## **1.2. Pertanyaan Penelitian**

Berdasarkan latar belakang yang ada, masalah yang dapat dirumuskan dalam penelitian ini sebagai berikut :

1. Bagaimana penerapan algoritma random forest dalam memprediksi penyakit liver berdasarkan fitur-fitur medis yang tersedia pada dataset kaggle?
2. Bagaimana performa algoritma random forest dalam memprediksi penyakit liver jika dievaluasi menggunakan metric evaluasi seperti accuracy, precision, reccal, f1 score, dan ROC-AUC?

## **1.3. Tujuan Penelitian**

Berdasarkan pertanyaan penelitian maka penelitian ini bertujuan untuk sebagai berikut:

1. Menjelaskan dan menerapkan algoritma Random Forest dalam memprediksi penyakit liver berdasarkan fitur-fitur medis yang terdapat pada dataset Kaggle.

2. Menganalisis performa algoritma Random Forest dalam memprediksi penyakit liver dengan menggunakan metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, *F1-score*, dan *ROC-AUC*.

#### 1.4. Kerangka Kerja Penelitian

Tabel 1.1 *Framework* Penelitian

Tahap	Teknik yang Digunakan	Tujuan
Data preparation	Imputasi, encoding	Menyiapkan data bersih dan siap model
Model training	RandomForestClassifier	Melatih model untuk mempelajari pola
Evaluasi performa	Accuracy, F1, AUC, confusion matrix	Mengukur kinerja model
Validasi model	Train-Test Split (80:20)	Menguji kemampuan generalisasi model dan menghindari overfitting
Feature importance	Analisis variabel penting	Mengetahui fitur paling berpengaruh