

**EVALUATION OF SMOTE TECHNIQUE IN THE
COMPARASION OF XGBOOST AND RANDOM FOREST
ALGORITHMS FOR LIVER DISEASE PREDICTION**

SKRIPSI NON REGULER JALUR SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



Disusun oleh :

WAHYUTRI NUR ROHMAN

22.11.5223

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**EVALUATION OF SMOTE TECHNIQUE IN THE COMPARASION
OF XGBOOST AND RANDOM FOREST ALGORITHMS
FOR LIVER DISEASE PREDICTION**

SKRIPSI NON REGULER JALUR SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



Disusun oleh :

WAHYUTRI NUR ROHMAN

22.11.5223

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

JALUR NON-REGULER

EVALUATION OF SMOTE TECHNIQUE IN THE COMPARASION
OF XGBOOST AND RANDOM FOREST ALGORITHMS
FOR LIVER DISEASE PREDICTION

yang disusun dan diajukan oleh

Wahyutri Nur Rohman

22.11.5123

telah disetujui oleh Dosen Pembimbing
pada tanggal 19 November 2025

Dosen Pembimbing,



I Made Artha Agastya, S.T., M.Eng, Ph.D.
NIK. 190302352

HALAMAN PENGESAHAN

JALUR NON-REGULER

EVALUATION OF SMOTE TECHNIQUE IN THE COMPARASION
OF XGBOOST AND RANDOM FOREST ALGORITHMS
FOR LIVER DISEASE PREDICTION

yang disusun dan diajukan oleh

Wahyutri Nur Rohman

22.11.5223

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 November 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Majid Rahardi, S.Kom., M.Eng
NIK. 190302393

Theophilus Bayu Sasongko, S.Kom, M.Eng
NIK. 190302375

I Made Artha Agastya, S.T., M.Eng, Ph.D
NIK. 190302352

Laporan ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 November 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriul, M.Kom
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Wahyutri Nur Rohman

NIM : 22.11.5223

Menyatakan bahwa Laporan dengan judul berikut:

Evaluation of SMOTE Technique in the Comparison of XGBoost and Random Forest Algorithms for Liver Disease Prediction

Dosen Pembimbing : I Made Artha Agastya, S.T., M.Eng, Ph.D.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan kegiatan SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidak-benaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 19 November 2025

Yang Menyatakan,



Wahyutri Nur Rohman

HALAMAN PERSEMBAHAN

Alhamdulillahirobbil'alamin puji syukur kehadiran Allah SWT. Dengan izin-Nya, penulis dapat menyelesaikan Tugas Akhir ini sebagai wujud tanggung jawab akademik sekaligus langkah akhir dalam menempuh pendidikan sarjana. Kehadiran karya ini menjadi bukti perjalanan panjang penuh doa, usaha, dan pengorbanan. Dengan penuh rasa syukur, penulis mempersembahkan karya ini kepada:

1. Dengan segenap rasa syukur, penulis haturkan puji kehadiran Allah SWT yang atas rahmat dan ridha-Nya telah memberikan kekuatan, kesehatan, dan kemudahan, sehingga penulis dapat menyelesaikan penelitian ini.
2. Ibu Sulastri dan Bapak Suyatno, kedua orang tua tercinta, yang menjadi sumber kekuatan penulis. Atas segala limpahan kasih sayang yang tak hingga, doa yang tiada henti, dukungan moril maupun material yang tak pernah putus, serta pengorbanan dalam mengusahakan segala hal demi menunjang kehidupan dan studi penulis.
3. Saudara kandung saya, Alfisyah Putri, yang selalu menjadi teman berdiskusi, penghibur di saat penat, serta penyemangat dalam setiap proses perjalanan ini.
4. Bapak I Made Artha Agastya, S.T., M.Eng, Ph.D., selaku Dosen Pembimbing Kolaborasi Riset, atas segala bimbingan, arahan, semangat yang diberikan, serta kesabarannya dalam membimbing penulis hingga terselesaikannya penelitian ini.
5. Kepada seseorang yang berarti, Ariska Tanzila Fatma, yang selalu menjadi penyemangat dalam kehidupan sehari-hari dan sumber motivasi ketika penulis kehilangan arah. Terima kasih atas perhatian, doa, dan semangat yang senantiasa diberikan.
6. Kepada keluarga besar kontrakan 42A, kakak tingkat dan teman seperjuangan yang telah menjadi tempat bernaung, berbagi cerita, tawa, serta semangat selama masa perkuliahan. Terima kasih atas kebersamaan dan kenangan berharga yang menjadi bagian dari perjalanan ini.

7. Kepada rekan-rekan Markas Besar Perantau. Alif, Andhika, Dhito, Dimas, Ferdy, Iksan, Kevan, Mirza, Nufus, Ulin, Yufana. terima kasih atas canda, dukungan, dan solidaritas yang selalu menghidupkan hari-hari selama perkuliahan.
8. Kepada teman-teman Discord, Abim, Alvi, Dayat, Hamdika, Hegar, Hekal, Huda, Pian, Royan, Satya, Tirta, Umam, Vano. Terima kasih atas tawa, candaan, serta diskusi malam yang tak terlupakan. Kebersamaan ini menjadi warna tersendiri yang menemani perjalanan akademik penulis hingga akhir.
9. Kepada teman-teman kost Alkindhi 62A, Afan, Anang, Angga, Arif, Ilham, Naufal, Ricky, Wahyu, dan Yazid. Terimakasih menjadi keluarga kedua selama masa perkuliahan, tempat berbagi tawa, keluh kesah, serta semangat dalam setiap langkah perjalanan ini.
10. Seluruh teman seperjuangan di kelas 22IF11 yang tidak dapat penulis sebutkan satu per satu terima kasih atas kebersamaan dan dukungan yang terjalin sejak awal perkuliahan hingga saat ini.
11. Wahyutri Nur Rohman, penulis penelitian ini, yang terus berusaha belajar, tumbuh, dan berproses dalam menghadapi setiap tantangan. Semoga karya ini menjadi langkah kecil menuju impian yang lebih besar serta memberikan manfaat bagi pembaca dan pengembangan ilmu pengetahuan.

KATA PENGANTAR

Segala Puji serta Syukur penulis panjatkan kehadirat Allah SWT, Tuhan Semesta Alam, yang karena rahmat dan karunia-Nya, penulis telah diberikan kemudahan dan kelancaran dalam menyelesaikan laporan kelulusan ini. Dalam menyelesaikan penelitian dan penyusunan laporan ini, penulis menyadari bahwa keberhasilan ini tidak akan terwujud tanpa adanya bimbingan, dukungan, dan bantuan tulus dari berbagai pihak. Oleh karena itu, pada kesempatan yang berharga ini, penulis ingin menyampaikan ucapan terima kasih yang mendalam kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusri, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom. selaku Ketua Program Studi S1 Informatika Universitas AMIKOM Yogyakarta.
4. Bapak I Made Artha Agastya, S.T., M.Eng, Ph.D. selaku Dosen Pembimbing penulis.
5. Ibu Sulastri dan Bapak Suyatno selaku orang tua penulis yang selalu memberikan cinta, kasih sayang, doa, dan pengorbanan yang sangat besar kepada penulis.

Penulis menyadari bahwa laporan ini masih jauh dari kata sempurna dan memiliki keterbatasan. Oleh karena itu, segala bentuk kritik dan saran yang membangun sangat diharapkan bagi perbaikan penelitian selanjutnya.

Yogyakarta, 6 November 2025

Penulis

DAFTAR ISI

Halaman Judul	1
Halaman Persetujuan	ii
Halaman Pengesahan	iii
Halaman Pernyataan Keaslian Karya	iv
Halaman Persembahan	v
Kata Pengantar	vii
Daftar Isi	viii
Daftar Tabel	x
Daftar Gambar	xi
Intisari	xii
Abstract	xiii
Bab I Pendahuluan	1
1.1. Gambaran Umum	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah	2
1.4. Tujuan	3
Bab II Tinjauan Pustaka	4
2.1. Studi Literatur	4
2.2. Landasan Teori	5
BAB III Metode Penelitian	6
3.1. Metode	6
3.1.1 Pengambilan Dataset	6
3.1.2 Exploratory Data Analysis (EDA)	7
3.1.3 Preprocessing Data	7
3.1.4 Split Data	8
3.1.5 Modelling	9
3.1.6 Evaluasi	10
BAB IV Pembahasan	14
4.1. Hasil	14

4.1.1	Pengambilan dan Analisa dataset	14
4.1.2	Exploratory Data Analysis (EDA)	15
4.1.3	Penanganan missing values	22
4.1.4	Encoding	23
4.1.5	Balancing data	24
4.1.6	Normalisasi data	26
4.1.7	Split data	27
4.1.8	Modelling	27
4.1.9	Evaluasi	27
BAB V	Kesimpulan	35
5.1.	Kesimpulan	35
5.2.	Saran	36
Referensi		37
Curriculum Vitae		40
Lampiran dan Bukti Pendukung		41
a.	Letter of Acceptance (LOA)	41
b.	Lembar Review	42
c.	Bukti Terbit/Terindex [wajib, sertakan tampilan bukti terbit]	47
d.	Bukti pembayaran	47

DAFTAR TABEL

Tabel 3. 1 Dataset.....	6
Tabel 3. 2 <i>Confusion Matrix</i>	11
Tabel 4. 1 Dataset	14
Tabel 4. 2 Missing Values.....	23
Tabel 4. 3 Penanganan Missing Values.....	23
Tabel 4. 4 Label Encoding.....	24
Tabel 4. 5 Normalisasi Data.	26
Tabel 4. 6 Split Data.....	27
Tabel 4. 7 Tabel Hasil Evaluasi tidak menggunakan Smote.....	31
Tabel 4. 8 Tabel Hasil Evaluasi menggunakan Smote.....	31
Tabel 4. 9 Performa Algoritma Xgboost.....	33
Tabel 4. 10 Performa Algoritma Random Forest.....	33



DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	6
Gambar 4. 1 Box Plot Age vs Result	16
Gambar 4. 2 Box Plot Total Bilirubin vs Result.	16
Gambar 4. 3 Box Plot Direct Bilirubin vs Result.	17
Gambar 4. 4 Box Plot Alkphos Alkaline Phosphotase vs Result.....	18
Gambar 4. 5 Box Plot Sgpt Alamine Aminotransferase vs Result.....	18
Gambar 4. 6 Box Plot Sgot Aspartate Aminotransferase vs Result.	19
Gambar 4. 7 Box Plot Total Protiens vs Result.....	20
Gambar 4. 8 Box Plot ALB Albumin vs Result.	21
Gambar 4. 9 Box Plot A/G Ratio vs Result.	21
Gambar 4. 10 Sebelum dilakukan Balancing data.	25
Gambar 4. 11 Setelah dilakukan Balancing data.	26
Gambar 4. 12 Extreme Gradient Boost Confusion Matrix without smote.	28
Gambar 4. 13 Random Forest Confusion Matrix without Smote.	29
Gambar 4. 14 Extreme Gradient Boost Confusion Matrix.....	30
Gambar 4. 15 Random Forest Confusion Matrix.....	31

INTISARI

Di banyak negara, termasuk Indonesia, penyakit hati tetap menjadi penyebab utama morbiditas dan mortalitas. Deteksi dini memainkan peran krusial dalam meningkatkan hasil pengobatan. Studi ini mengevaluasi kinerja dua model pembelajaran mesin yang luas digunakan, Random Forest dan XGBoost, untuk memprediksi penyakit hati, dengan menggunakan teknik penyeimbangan SMOTE untuk mengatasi ketidakseimbangan kelas. Tujuan utama adalah untuk meningkatkan keadilan model, mengurangi *overfitting*, dan meningkatkan sensitivitas terhadap kelas minoritas. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Model XGBoost mencapai akurasi rata-rata 99,53%, presisi 99,46%, recall 98,89%, dan F1-score sebesar 99,18%, sementara model Random Forest mencapai akurasi rata-rata 99,83%, presisi 99,84%, recall 99,58%, dan F1-score 99,71%. Kedua model menunjukkan kemampuan prediktif yang sangat baik, dengan Random Forest sedikit lebih unggul daripada XGBoost. Hasil ini menyoroti pentingnya penyeimbangan data dan validasi model yang kuat dalam mengembangkan model pembelajaran mesin yang andal untuk pengambilan keputusan di bidang kesehatan.

Kata kunci: Data Mining, Prediksi Penyakit Hati, Random Forest, XGBoost, Evaluasi Model

ABSTRACT

In many countries, including Indonesia, liver disease remains a major cause of morbidity and mortality. Early detection plays a crucial role in improving treatment outcomes. This study evaluates the performance of two widely used machine learning models Random Forest and XGBoost for predicting liver disease, employing the SMOTE balancing technique to address class imbalance. The primary objectives are to enhance model fairness, reduce overfitting, and improve sensitivity toward the minority class. Model performance is assessed using accuracy, precision, recall, and F1-score metrics. The XGBoost model achieved an average accuracy of 99.53%, precision of 99.46%, recall of 98.89%, and F1-score of 99.18%, while the Random Forest model attained an average accuracy of 99.83%, precision of 99.84%, recall of 99.58%, and F1-score of 99.71%. Both models demonstrated excellent predictive capability, with Random Forest slightly outperforming XGBoost. These results highlight the importance of data balancing and robust model validation in developing reliable machine learning models for healthcare decision-making.

Keyword: Data Mining, Liver Disease Prediction, RandomForest, XGBoost, Model Evaluation