



**IMPLEMENTASI IMPUTASI MISSFOREST PADA
KLASIFIKASI PENYAKIT SIROSIS HATI MENGGUNAKAN
SMOTE DENGAN METODE *XGBOOST* DAN *RANDOM FOREST***

Skripsi

**Untuk memenuhi persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

CHRISTIAN ALDI JEDUTHUN

1911016210016

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JULI 2026

SKRIPSI

IMPLEMENTASI IMPUTASI MISSFOREST PADA KLASIFIKASI PENYAKIT SIROSIS HATI MENGGUNAKAN SMOTE DENGAN METODE XGBOOST DAN RANDOM FOREST

Oleh :

CHRISTIAN ALDI JEDUTHUN

NIM. 1911016210016

telah dipertahankan di depan Dosen Penguji pada tanggal 16 Juli 2026

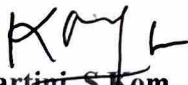
Susunan Penguji :

Pembimbing Utama



Triando H. Saragih, S.Kom., M.Kom.
NIP. 199308242019031012

Pembimbing Pendamping



Dwi Kartini, S.Kom., M.Kom.
NIP. 198704212012122003

Penguji

1.



Fatma Indriani, S.T., M.I.T., Ph.D
NIP. 198404202008122004

2



Rudy Hertono, S.Kom., M.Kom
NIP. 198809252022031003

Banjarbaru, 27 Juli 2026

Koordinator Program Studi Ilmu Komputer




Dwi Kartini, S.Kom., M.Kom.
NIP. 198704212012122003

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarmasin, 15 Juli 2026

Yang Menyatakan,

A handwritten signature in black ink, consisting of a large, stylized 'C' followed by a series of loops and a long horizontal stroke extending to the right.

Christian Aldi Jeduthun

NIM. 1911016210016

ABSTRAK

IMPLEMENTASI IMPUTASI MISSFOREST PADA KLASIFIKASI PENYAKIT SIROSIS HATI MENGGUNAKAN SMOTE DENGAN METODE *XGBOOST* DAN *RANDOM FOREST*

(Oleh : Christian Aldi Jeduthun; Pembimbing: Triando Hamonangan Saragih, S.Kom., M.Kom. dan Dwi Kartini, S.Kom., M.Kom; 2026; 74 halaman)

Sirosis hati adalah penyakit kronis yang menyerang hati dan membutuhkan deteksi dini untuk mencegah komplikasi lebih lanjut. Namun, data medis seringkali memiliki tantangan berupa nilai hilang dan ketidakseimbangan kelas antara pasien positif dan negatif. Penelitian ini mengusulkan kerangka kerja klasifikasi dengan menerapkan imputasi missForest untuk menangani data hilang secara non-parametrik dan teknik oversampling *Synthetic Minority Over-sampling Technique* (SMOTE) digunakan untuk menyeimbangkan kelas. Setelah pra-pemrosesan, data diklasifikasikan menggunakan dua algoritma, yaitu Extreme Gradient Boosting (XGBoost) dan Random Forest, yang dipilih karena kinerjanya yang baik pada data tabular medis. Evaluasi model dilakukan menggunakan akurasi, presisi, recall, dan AUC, dengan penekanan utama pada nilai recall untuk memastikan deteksi pasien positif yang optimal. Hasil penelitian menunjukkan bahwa model Random Forest dengan SMOTE pada skema pembagian data 60:40 mencapai nilai recall tertinggi sebesar 65,09%. Meskipun akurasi model ini sebesar 66,66%, peningkatan recall yang signifikan menunjukkan kemampuan model yang lebih baik dalam mengidentifikasi pasien sirosis hati dibandingkan model tanpa SMOTE. Sementara itu, model XGBoost tanpa SMOTE pada skema 90:10 mencapai akurasi tertinggi 88,09% namun memiliki recall lebih rendah. Kesimpulan penelitian ini adalah bahwa kombinasi imputasi missForest dan SMOTE secara signifikan meningkatkan kemampuan model dalam mendeteksi pasien sirosis hati. Temuan ini menunjukkan bahwa pendekatan yang diusulkan berpotensi untuk dikembangkan sebagai sistem pendukung keputusan dalam diagnosis dini sirosis hati berbasis data.

Kata Kunci: MissForest, *XGBoost*, *Random Forest*, SMOTE

ABSTRACT

IMPLEMENTATION OF MISSFOREST IMPUTATION IN CLASSIFICATION OF LIVER CIRRHOSIS DISEASE USING SMOTE WITH XGBOOST AND RANDOM FOREST METHOD

(Oleh : Christian Aldi Jeduthun; Pembimbing: Triando Hamonangan Saragih, S.Kom., M.Kom. dan Dwi Kartini, S.Kom., M.Kom; 2026; 74 pages)

Liver cirrhosis is a chronic disease that affects the liver and requires early detection to prevent further complications. However, medical data often presents challenges in the form of missing values and class imbalances between positive and negative patients. This study proposes a classification framework that applies missForest imputation to handle missing data non-parametrically and uses the Synthetic Minority Over-sampling Technique (SMOTE) to balance the classes. After preprocessing, the data were classified using two algorithms—Extreme Gradient Boosting (XGBoost) and Random Forest—which were selected for their strong performance on tabular medical data. Model evaluation was conducted using accuracy, precision, recall, and AUC, with a primary emphasis on recall to ensure optimal detection of positive patients. The results show that the Random Forest model with SMOTE, using a 60:40 data split, achieved the highest recall value of 65.09%. Although this model's accuracy was 66.66%, the significant increase in recall indicates that the model is better at identifying patients with liver cirrhosis compared to the model without SMOTE. Meanwhile, the XGBoost model without SMOTE on the 90:10 split achieved the highest accuracy of 88.09% but had a lower recall. The conclusion of this study is that the combination of missForest imputation and SMOTE significantly improves the model's ability to detect patients with liver cirrhosis. These findings suggest that the proposed approach has the potential to be developed as a data-driven decision support system for the early diagnosis of liver cirrhosis.

Kata Kunci: *MissForest, XGBoost, Random Forest, SMOTE*

PRAKATA

Puji syukur penulis panjatkan kepada Tuhan kita Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “IMPLEMENTASI IMPUTASI MISSFOREST PADA KLASIFIKASI PENYAKIT SIROSIS HATI MENGGUNAKAN SMOTE DENGAN METODE XGBOOST DAN RANDOM FOREST” untuk memenuhi syarat dalam menyelesaikan pendidikan program Srata-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

1. Tuhan Yang Maha Esa, karena atas kasih karunia-Nya lah penulis dapat menyelesaikan skripsi ini.
2. Bapak David Aris Raharja, Ibu Hastari Utami, adik Yosia Gracenada Renandita, serta keluarga atas doa dan dukungan dalam menyelesaikan skripsi ini.
3. Bapak Triando Hamonangan Saragih, S.Kom., M.Kom. selaku dosen pembimbing utama dan dosen pembimbing akademik yang meluangkan waktu untuk membantu demi kelancaran dalam menyelesaikan skripsi ini. Atas izin dari beliau jugalah skripsi ini dapat diselesaikan.
4. Ibu Dwi Kartini, S.Kom, M.Kom. selaku dosen pembimbing pendamping yang turut membantu serta meluangkan waktu dalam kelancaran menyelesaikan skripsi ini.
5. Ibu Fatma Indriani, S.T., M.I.T., Ph.D dan Bapak Rudy Herteno, S.Kom., M.Kom. selaku dosen ketua penguji dan anggota penguji dalam skripsi ini.
6. Seluruh Dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
7. Arsenal sebagai tim sepak bola yang selalu menemani penulis selagi dalam proses pembuatan skripsi.

8. Ucapan terima kasih kepada Alvis Jauza Prisma, Muhammad Bayu Ariyadi, Rayhan Surya Ibrahim, Mita Fauziyah, Dyna Ariva Maulidya, Tiara Nabaneeta Asaahi, Mochammad Sobar, Anita Absyari dan Naufal Nurmuzakki selaku sahabat penulis yang memberikan dukungan dalam proses mengerjakan skripsi ini.
9. Seluruh keluarga Ilmu Komputer angkatan 2019 yang memberikan dukungan dari awal sampai akhir.
10. Seluruh keluarga team Study *Qdenga and Dengue Hospitalization* selaku team yang memberikan dukungan dari awal sampai akhir.
11. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan serupa berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya.

Banjarmasin, 16 Juli 2026

Christian Aldi Jeduthun

DAFTAR ISI

	Halaman
PERNYATAAN.....	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
BAB I	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	5
1.5 Batasan Masalah	5
BAB II.....	6
2.1 Kajian terdahulu.....	6
2.2 Landasan Teori.....	12
2.2.1 Penyakit Sirosis Hati	12
2.2.2 Data Mining.....	12
2.2.3 <i>Preprocessing</i>	13
2.2.4 <i>Missing Value</i>	13
2.2.5 MissForest	13
2.2.6 MinMaxScaler	15
2.2.7 Klasifikasi.....	16

2.2.8	Random Forest.....	17
2.2.9	Boosting.....	18
2.2.10	Extreme Gradient Boost (XGBoost)	18
2.2.11	Confusion Matrix.....	19
2.2.12	Area Under Curve (AUC)	20
BAB III		22
3.1	Alat Penelitian.....	22
3.2	Bahan Penelitian	22
3.3	Variabel Penelitian.....	23
3.4	Prosedur Penelitian	23
BAB IV		27
4.1	Hasil Penelitian	27
4.1.1	Pengumpulan Data.....	27
4.1.2	Pra Proses Data.....	31
4.1.2.1	Imputasi Missing Value.....	31
4.1.2.2	Preprocessing.....	35
4.1.3	Pembagian Data.....	37
4.1.4	Penyeimbangan Data	37
4.1.5	Klasifikasi.....	38
4.1.5.1	XGBoost	39
4.1.5.2	Random Forest.....	49
4.1.6	Evaluasi	58
4.2	Pembahasan.....	59
BAB V PENUTUP.....		68
5.1	Kesimpulan	68
5.2	Saran	69
DAFTAR PUSTAKA		70
LAMPIRAN.....		1

DAFTAR TABEL

Tabel 1 Keaslian Penelitian.....	8
Tabel 2 Rancangan Penelitian.....	11
Tabel 3 Tabel Confusion Matrix	19
Tabel 4 Evaluasi nilai AUC	20
Tabel 5 Deskripsi data.....	27
Tabel 6 Pembagian distribusi data pasien penyakit sirosis	29
Tabel 7 karakteristik data	30
Tabel 8 Contoh data Cirrhosis Prediction	31
Tabel 9 Jumlah data setelah proses <i>dropNA</i>	32
Tabel 10 Data tabel urutan jumlah missing value.....	32
Tabel 11 Contoh tabel data Cirrhosis.....	33
Tabel 12 Urutan pengerjaan missForest.....	33
Tabel 13 Tabel pengisian nilai fitur Prothrombin menggunakan Random Forest	33
Tabel 14 Pengisian nilai NA pada fitur Cholesterol menggunakan nilai fitur sebelumnya.....	34
Tabel 15 Pengisian nilai NA pada fitur Tryglicerides menggunakan nilai fitur sebelumnya.....	34
Tabel 16 hasil imputasi MissForest	35
Tabel 17 Data hasil Preprocessing	35
Tabel 18 Data hasil MinMax Normalization	36
Tabel 19 Pembagian dataset pada data dropNA	37
Tabel 20 Pembagian dataset pada data imputasi missForest	37
Tabel 21 Tabel data setelah diseimbangkan menggunakan SMOTE pada split data 60:40.....	38
Tabel 22 Hasil Kinerja Parameter max_depth tanpa menggunakan SMOTE pada split data 60:40	39
Tabel 23 Hasil Kinerja Parameter max_depth tanpa menggunakan SMOTE pada split data 70:30	40

Tabel 24 Hasil Kinerja Parameter max_depth tanpa menggunakan SMOTE pada split data 80:20	40
Tabel 25 Hasil Kinerja Parameter max_depth tanpa menggunakan SMOTE pada split data 90:10	41
Tabel 26 Hasil Kinerja XGboost SMOTE pada parameter terbaik dari Manual Grid pada split data non SMOTE.....	41
Tabel 27 Hasil Kinerja XGBoost non imputasi pada parameter terbaik dari split data nonSMOTE	42
Tabel 28 Hasil Kinerja XGBoost SMOTE non imputasi pada parameter terbaik dari split data nonSMOTE	42
Tabel 29 Hasil kinerja parameter learning_rate pada split data 60:40	43
Tabel 30 Hasil kinerja parameter learning_rate pada split data 70:30	43
Tabel 31 Hasil kinerja parameter learning_rate pada split data 80:20	44
Tabel 32 Hasil kinerja parameter learning_rate pada split data 90:10	44
Tabel 33 Kinerja XGBoost SMOTE pada parameter terbaik di masing-masing Split data Non SMOTE	45
Tabel 34 Kinerja XGBoost non imputasi pada parameter terbaik	45
Tabel 35 Kinerja XGBoost SMOTE non imputasi pada parameter terbaik.....	45
Tabel 36 Hasil kinerja parameter n_estimator pada split data 60:40	46
Tabel 37 Hasil kinerja parameter n_estimator pada split data 70:30	46
Tabel 38 Hasil kinerja parameter n_estimator pada split data 80:20	47
Tabel 39 Hasil kinerja parameter n_estimator pada split data 90:10	47
Tabel 40 Hasil Kinerja XGBoost SMOTE dengan parameter terbaik di split data NonSMOTE	48
Tabel 41 Hasil Kinerja XGBoost non imputasi pada parameter terbaik.....	48
Tabel 42 Hasil Kinerja XGBoost SMOTE non imputasi pada parameter terbaik	48
Tabel 43 Hasil kinerja parameter Max_depth pada split data 60:40.....	49
Tabel 44 Hasil kinerja parameter Max_depth pada split data 70:30.....	50
Tabel 45 Hasil kinerja parameter Max_depth pada split data 80:20.....	51
Tabel 46 grafik kinerja parameter Max_depth pada split data 90:10.....	52

Tabel 47 Hasil Kinerja Random Forest SMOTE pada parameter terbaik non SMOTE	52
Tabel 48 Hasil Kinerja Random Forest non Imputasi pada parameter terbaik ..	53
Tabel 49 Hasil Kinerja Random Forest SMOTE non imputasi pada parameter terbaik.....	53
Tabel 50 hasil kinerja parameter N_estimator pada split data 60:40	53
Tabel 51 Hasil kinerja parameter N_estimator pada split data 70:30	54
Tabel 52 Hasil kinerja parameter N_estimator pada split data 80:20	56
Tabel 53 Hasil kinerja parameter N_estimator pada split data 90:10	56
Tabel 54 Hasil Kinerja Random Forest SMOTE pada parameter terbaik nonSMOTE	57
Tabel 55 Hasil Kinerja Random Forest non Imputasi pada parameter terbaik ..	57
Tabel 56 Hasil Kinerja Random Forest SMOTE non Imputasi pada parameter terbaik.....	57
Tabel 57 contoh confusion matrix XGBoost pada split data 60:40	58
Tabel 58 Tabel hasil kinerja keseluruhan nonSMOTE pada imputasi missForest dan DropNA.....	60
Tabel 59 Hasil kinerja keseluruhan Perbandingan SMOTE pada imputasi missForest dan dropNA.....	61

DAFTAR GAMBAR

Gambar 1 Alur <i>MinMaxScaler</i>	16
Gambar 2 Pohon Keputusan.....	17
Gambar 3 Alur Bagging (<i>Bootstrap Aggregating</i>)	18
Gambar 4 Alur Penelitian.....	24
Gambar 5 Grafik kinerja akurasi pada data non SMOTE	62
Gambar 6 Grafik kinerja <i>Precision</i> pada data non SMOTE	62
Gambar 7 Grafik kinerja <i>Recall</i> pada data non SMOTE	63
Gambar 8 Perbandingan Kinerja AUC pada data non SMOTE.....	63
Gambar 9 Perbandingan Kinerja akurasi pada data yang dilakukan SMOTE ...	64
Gambar 10 Perbandingan Kinerja presisi pada data yang dilakukan SMOTE ..	64
Gambar 11 Perbandingan Kinerja <i>recall</i> pada data yang dilakukan SMOTE ...	65
Gambar 12 Perbandingan Kinerja AUC pada data yang dilakukan SMOTE	65

DAFTAR LAMPIRAN

Lampiran

Lampiran 1 *Source Code MissForest* menggunakan Rstudio

Lampiran 2 *Source Code* melakukan Evaluasi kinerja metode *XGBoost* menggunakan *Python*

Lampiran 3 *Source Code* melakukan Evaluasi kinerja metode *XGBoost* dan penambahan SMOTE menggunakan *Python*

Lampiran 4 *Source Code* melakukan Evaluasi kinerja metode *Random Forest* menggunakan *Python*

Lampiran 5 *Source Code* melakukan Evaluasi kinerja metode *Random Forest* dan penambahan SMOTE menggunakan *Python*

Lampiran 6 Riwayat Hidup