



**Penerapan *Local Interpretable Model-agnostic Explanations*
(LIME) Dalam Model *Gradient Boosting* untuk Klasifikasi
Kondisi Pulang Pasien RSUD Ulin Banjarmasin**

SKRIPSI

**untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
YUSUF FADILLAH
NIM. 2211017130003**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2026**



**Penerapan *Local Interpretable Model-agnostic Explanations*
(LIME) Dalam Model *Gradient Boosting* untuk Klasifikasi
Kondisi Pulang Pasien RSUD Ulin Banjarmasin**

SKRIPSI

**untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
YUSUF FADILLAH
NIM. 2211017130003**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2026**

SKRIPSI

Penerapan *Local Interpretable Model-agnostic Explanations* (LIME) Dalam Model *Gradient Boosting* untuk Klasifikasi Kondisi Pulang Pasien RSUD Ulin Banjarmasin

Oleh
YUSUF FADILLAH
NIM. 2211017310003

Telah dipertahankan pada hari Kamis tanggal 4 Juni 2026 dan telah disetujui
oleh dosen pembimbing dan dosen penguji sebagai berikut:

Pembimbing I



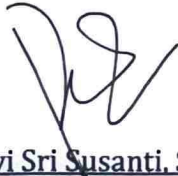
Selvi Annisa, S.Si., M.Si
NIP. 199212262022032016

Penguji I



Rifqi Aulya Rahman, S.Mat., M.Si
NIP. 199708072024061003

Pembimbing II



Dewi Sri Susanti, S.Si., M.Si
NIP. 197305161999032002

Penguji II



dr. Siti Ningsih, M.Kes
NIP. 198204192011012004

Banjarbaru, 10 Juni 2026

Mengetahui,
Koordinator Program Studi Statistika
FMIPA ULM




Dewi Anggraini, S.Si., M.App.Sci., Ph.D.
NIP. 198303282005012001

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.

Banjarbaru, 10 Juni 2026



Yusuf Fadillah
NIM. 2211017310003

ABSTRAK

Penerapan *Local Interpretable Model-agnostic Explanations* (LIME) Dalam Model *Gradient Boosting* untuk Klasifikasi Kondisi Pulang Pasien RSUD Ulin Banjarmasin (Oleh: Yusuf Fadillah; Pembimbing: Selvi Annisa, S.Si., M.Si dan Dewi Sri Susanti, S.Si., M.Si, 2026; 54 halaman)

Tingginya angka mortalitas di Indonesia menunjukkan perlunya penguatan sistem kesehatan, khususnya dalam mendeteksi risiko kematian pasien secara lebih cepat dan tepat. Pemanfaatan data rekam medis melalui pendekatan *machine learning* dapat membantu klasifikasi kondisi pulang pasien berdasarkan faktor-faktor klinis dan administratif. Oleh karena itu, penelitian ini bertujuan membangun model klasifikasi kondisi pulang pasien di RSUD Ulin Banjarmasin menggunakan *Gradient Boosting Machine* (GBM) serta menginterpretasikan hasil prediksi menggunakan *Local Interpretable Model-agnostic Explanations* (LIME). Tahapan penelitian meliputi *preprocessing* data, pembagian data menjadi data *training* dan *testing*, penerapan *Synthetic Minority Oversampling Technique–Nominal Continuous* (SMOTE-NC), serta penentuan parameter optimal menggunakan *GridSearchCV*. Berdasarkan hasil perbandingan model, penerapan SMOTE-NC pada proporsi pembagian data 70:30 menghasilkan performa terbaik dengan nilai *sensitivity* sebesar 0.971, *specificity* sebesar 0.311, dan *F1-macro* sebesar 0.663 pada data *testing*. Interpretasi model menggunakan LIME menunjukkan nilai LOS yang lebih tinggi, umur pasien yang lebih besar, serta jumlah penyakit sekunder yang lebih banyak cenderung meningkatkan probabilitas klasifikasi pasien masuk ke kelas meninggal. Sebaliknya, jumlah tindakan, kategori penyakit tertentu, serta jumlah penyakit sekunder yang lebih sedikit cenderung meningkatkan probabilitas klasifikasi pasien masuk ke kelas membaik/sembuh. Hasil penelitian menunjukkan bahwa kombinasi GBM dan SMOTE-NC mampu menghasilkan performa klasifikasi yang baik, sedangkan LIME dapat meningkatkan interpretabilitas model dengan menjelaskan kontribusi masing-masing variabel terhadap prediksi secara lokal. Penelitian ini diharapkan dapat mendukung deteksi dini risiko kematian pasien serta membantu pengambilan keputusan klinis dan evaluasi pelayanan kesehatan di RSUD Ulin Banjarmasin.

Kata kunci: *Gradient Boosting Machine*, SMOTE-NC, *Local Interpretable Model-agnostic Explanations*, Klasifikasi, Rekam Medis, Kondisi Pulang Pasien.

ABSTRACT

Application of Local Interpretable Model-agnostic Explanations (LIME) in the Gradient Boosting Model for Classification of Patient Discharge Status at Ulin Regional General Hospital Banjarmasin (By: Yusuf Fadillah; Supervisors: Selvi Annisa, S.Si., M.Si. and Dewi Sri Susanti, S.Si., M.Si., 2026; 54 pages)

The high mortality rate in Indonesia indicates the need to strengthen the healthcare system, particularly in detecting patient mortality risk more quickly and accurately. The utilization of medical record data through a machine learning approach can assist in classifying patient discharge status based on clinical and administrative factors. Therefore, this study aims to develop a patient discharge status classification model at Ulin Regional General Hospital Banjarmasin using the Gradient Boosting Machine (GBM) algorithm and to interpret the prediction results using Local Interpretable Model-agnostic Explanations (LIME). The research stages included data preprocessing, splitting the data into training and testing sets, applying the Synthetic Minority Oversampling Technique–Nominal Continuous (SMOTE-NC), and determining optimal parameters using GridSearchCV. Based on the model comparison results, the implementation of SMOTE-NC with a 70:30 data split proportion produced the best performance, with a sensitivity value of 0.971, specificity of 0.311, and F1-macro of 0.663 on the testing data. The LIME interpretation showed that higher Length of Stay (LOS), older patient age, and a greater number of secondary diseases tended to increase the probability of patients being classified into the deceased class. In contrast, the number of medical procedures, certain disease categories, and fewer secondary diseases tended to increase the probability of patients being classified into the recovered/improved class. The results indicate that the combination of GBM and SMOTE-NC is capable of producing good classification performance, while LIME improves model interpretability by explaining the contribution of each variable to local predictions. This study is expected to support early detection of patient mortality risk and assist clinical decision-making as well as healthcare service evaluation at Ulin Regional General Hospital Banjarmasin.

Keywords: Gradient Boosting Machine, SMOTE-NC, Local Interpretable Model-agnostic Explanations, Classification, Medical Records, Patient Discharge Status.

PRAKATA

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Penerapan *Local Interpretable Model-agnostic Explanations* (LIME) dalam Model *Gradient Boosting* untuk Klasifikasi Kondisi Pulang Pasien RSUD Ulin Banjarmasin”. Tugas akhir ini disusun sebagai salah satu syarat untuk memenuhi salah satu dalam menyelesaikan program sarjana Strata-1 (S-1) pada Program Studi Statistika FMIPA ULM. Penulis menyadari bahwa penyusunan skripsi ini tidak terlepas dari bantuan, dukungan, dan bimbingan berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Kedua orang tua dan keluarga yang selalu memberikan doa, dukungan, serta motivasi kepada penulis.
2. Ibu Selvi Annisa, S.Si., M.Si. dan Ibu Dewi Sri Susanti, S.Si., M.Si. selaku dosen pembimbing yang telah memberikan bimbingan, arahan, saran, dan masukan selama penyusunan skripsi.
3. Bapak Rifqi Aulya Rahman, S.Mat., M.Si. dan Ibu dr. Siti Ningsih, M.Kes. selaku dosen penguji yang telah memberikan kritik, saran, dan masukan demi penyempurnaan skripsi ini.
4. Seluruh dosen dan staf Program Studi Statistika yang telah memberikan ilmu dan bantuan selama masa perkuliahan.
5. Pihak RSUD Ulin Banjarmasin yang telah memberikan izin dan dukungan dalam proses pengambilan data penelitian.
6. Teman-teman *Funtastic* yang telah membantu dan memberikan semangat selama penyusunan skripsi ini.
7. Serta semua pihak yang yang tidak bisa penulis sebutkan satu per satu, tetapi kontribusinya memberikan makna tersendiri bagi penulis.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi perbaikan di masa mendatang. Penulis berharap skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang statistika dan penerapan machine learning dalam bidang kesehatan.

Banjarbaru, 4 Juni 2026

Yusuf Fadillah

DAFTAR ISI

PERNYATAAN.....	iii
ABSTRAK.....	iv
ABSTRACT.....	v
PRAKATA.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR LAMPIRAN	xi
DAFTAR LAMBANG DAN SINGKATAN.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA	5
2.1 Kajian Penelitian Terdahulu	5
2.2 Kajian Teori	7
2.2.1 Statistika Deskriptif.....	7
2.2.2 Penanganan <i>Imbalance</i> pada Data.....	10
2.2.3 Klasifikasi	11
2.2.4 <i>Confusion Matrix</i>	17
2.2.5 <i>Local Interpretable Model-agnostic Explanations</i> (LIME).....	18
2.2.6 Rekam Medis	20
2.2.7 <i>The International Classification of Diseases</i> (ICD).....	21
BAB III METODE PENELITIAN.....	23
3.1 Sumber Data.....	23
3.2 Variabel Penelitian.....	23
3.3 Prosedur Penelitian.....	25
BAB IV HASIL DAN PEMBAHASAN.....	29
4.1 <i>Pre-Processing</i>	29
4.2 Eksplorasi Data	31
4.2.1 Deskriptif Statistik Variabel Kategorik dengan Kondisi Pulang Pasien	31
4.2.2 Deskriptif Statistik Variabel Numerik dengan Kondisi Pulang Pasien	36
4.3 Pembagian Data <i>Traning</i> dan Data <i>Testing</i>	37
4.4 Penanganan <i>Imbalance</i> data	38
4.4.1 Menghitung Nilai Median dari Standar Deviasi Variabel Numerik	

4.4.2 Menghitung Jarak Eucledean	38
4.4.3 Membuat Data Sintetis	39
4.5 Pembangunan dan Pelatihan Model GBM	41
4.5.1 Membangun Awal Model	41
4.5.2 Pembuatan CART	42
4.5.3 Penambahan Nilai Pembaharuan	44
4.5.4 Menghitung Nilai <i>Log-odds</i> Akhir	46
4.5.5 Penentuan Proporsi Data <i>Training</i> dan <i>Testing</i>	46
4.5.6 Hasil <i>Tuning</i> Parameter	47
4.6 Pengujian Model GBM Data <i>Testing</i>	48
4.7 Penerapan Teknik LIME pada Model GBM	49
4.7.1 Membuat Sampel Tiruan untuk LIME	49
4.7.2 Menghitung Jarak dan Bobot	50
4.7.3 Pelatihan Model Lokal	51
4.7.4 Interpretasi Model GBM Menggunakan LIME	52
BAB V PENUTUP	54
5.1 Kesimpulan	54
5.2 Saran	54
DAFTAR PUSTAKA	56
LAMPIRAN	61
RIWAYAT HIDUP	88

DAFTAR GAMBAR

Gambar 1.1 Angka Kematian Kasar di Indonesia Tahun 2000-2023	1
Gambar 2.1 Statistika Deskriptif dengan Diagram Batang dan Lingkaran	9
Gambar 2.2 Ilustrasi <i>Decision Tree</i>	12
Gambar 2.3 Ilustrasi Teknik <i>Boosting</i>	13
Gambar 2.4 Ilustrasi <i>Cross Validation</i>	16
Gambar 2.5 Visualisasi Eksplanasi Model Menggunakan LIME	20
Gambar 3.1 Prosedur Analisis	28
Gambar 4.1 Proporsi Kondisi Pulang Pasien berdasarkan Jenis Ruangan	32
Gambar 4.2 Proporsi Kondisi Pulang Pasien berdasarkan Op dan Non OP	33
Gambar 4.3 Proporsi Kondisi Pulang Pasien berdasarkan Jumlah Penyakit Sekunder	33
Gambar 4.4 Proporsi Kondisi Pulang Pasien berdasarkan Jumlah Tindakan	34
Gambar 4.5 Proporsi Kondisi Pulang Pasien berdasarkan Kategori ICD-10	35
Gambar 4.6 Sebaran Umur berdasarkan Kondisi Pulang Pasien	36
Gambar 4.7 Sebaran LOS berdasarkan Kondisi Pulang Pasien	37
Gambar 4.8 Pohon CART Pertama <i>Gradient Boosting</i>	44
Gambar 4.9 Interpretasi Lokal Pasien ke-2 Menggunakan Teknik LIME	53
Gambar 4.10 Interpretasi Lokal Pasien ke-12 Menggunakan Teknik LIME	54

DAFTAR TABEL

Tabel 2.1 Kajian Penelitian Terdahulu	5
Tabel 2.2 Bentuk Tabel <i>Confusion Matrix</i>	17
Tabel 2.3 Kategori Dasar ICD-10.....	21
Tabel 3.1 Variabel Penelitian	23
Tabel 4.1 Penyederhanaan Variabel	29
Tabel 4.2 Proporsi Kelas Data Variabel Target	31
Tabel 4.3 Pembagian Data	37
Tabel 4.4 Contoh Data Kategorik Untuk Menghitung Jarak <i>Euclidean</i>	39
Tabel 4.5 Contoh Data Minoritas (Kelas Meninggal)	39
Tabel 4.6 Contoh Data untuk Membuat Data Sintetis.....	40
Tabel 4.7 Proporsi Data Training Sebelum dan Sesudah <i>Resampling</i>	40
Tabel 4.8 Daftar <i>Hyperparameter</i> dan Nilai <i>Grid Search</i> Model GBM	41
Tabel 4.9 Nilai Pembaharuan dan <i>Pseudo-Respons</i> Data Ke-1.....	47
Tabel 4.10 Kinerja GBM pada Berbagai Rasio Data dan SMOTE-NC.....	48
Tabel 4.11 Performa Model <i>Training</i> Setelah <i>Hyperparameter Tuning</i>	49
Tabel 4.12 <i>Confusion Matrix</i> Model GBM	49
Tabel 4.13 Sampel Tiruan (<i>z</i>) LIME Pasien Ke-1.....	51
Tabel 4.14 Nilai Kontribusi Setip Variabel.	53

DAFTAR LAMPIRAN

Lampiran 1. Data Penelitian	63
Lampiran 2. Data Setelah <i>Preprocessing</i> dan Penanganan <i>Imbalance</i>	64
Lampiran 3. Daftar Kombinasi <i>Hyperparameter Tuning</i>	65
Lampiran 4. Hasil Representasi Interpretabel Data Asli Dan Baru Pada Pasien Ke-1	66
Lampiran 5. Jarak dan Bobot Kedekatan Sampel Tiruan terhadap Data Asli Pada Pasien Ke-1	67
Lampiran 6. Nilai Kontribusi LIME pada Setiap prediktor untuk kelas "Membaik/Sembuh"	68
Lampiran 7. Nilai Kontribusi LIME pada Setiap prediktor untuk kelas "Meninggal"	68
Lampiran 8. <i>Syntax Python</i> untuk <i>Preprocessing</i> Data	70
Lampiran 9. <i>Syntax Python</i> Untuk Statistika Deskriptif.....	75
Lampiran 10. <i>Syntax Python</i> Untuk Pelatihan Model <i>Gradient Boosting</i> dengan Parameter <i>Default</i> Menggunakan Skema Tanpa SMOTE-NC Pada Rasio 70%, 80%, dan 90%.....	77
Lampiran 11. <i>Syntax Python</i> untuk Pelatihan Model <i>Gradient Boosting</i> dengan Parameter <i>Default</i> Menggunakan Skema SMOTE-NC Pada Rasio 70%, 80%, dan 90%.	81
Lampiran 12. <i>Syntax Python</i> untuk <i>Tuning Hyperparameter</i> Model dengan SMOTE-NC Rasio 70:30	85
Lampiran 13. <i>Syntax Python</i> untuk Nilai LIME.....	89

DAFTAR LAMBANG DAN SINGKATAN

a	Data pengamatan dari kelas minoritas
b	Data pengamatan yang terdekat dari minoritas
$d(a_i, b_i)$	Jarak antara dua nilai pengamatan pada prediktor ke- i
h	Jumlah variabel prediktor kategorik
q	Jumlah variabel prediktor numerik
a_{syn}	Data sintetis yang dihasilkan SMOTE-NC
a_{knn}	Data kelas minoritas yang terdekat dengan a_i
r	Bilangan random dari 0 hingga 1
k	Jumlah kelas pada variabel target
p	Proporsi kelas dari variabel target
A	Himpunan dalam suatu Node
o	Jumlah <i>subset</i> setelah <i>split</i>
n	Jumlah total data sebelum split
n_t	Jumlah data pada subset ke- t
$F_0(x)$	Model awal GBM
$F_m(x)$	Model pada iterasi ke- m
T_m	<i>Weak learner</i> pada iterasi ke- m
γ	Nilai prediksi model
v	<i>Learning rate</i>
L	Fungsi kerugian (<i>loss function</i>)
x	Variabel prediktor
y	Variabel target
$\tilde{y}_{im}$	Variabel target sementara ke- i iterasi ke- m
f	Model kompleks (<i>black box</i>)
g	Model lokal
z	Variabel prediktor yang di pertubasi
$\pi_x(z)$	Fungsi pembobotan lokal
$\Omega(g)$	Kompleksitas model lokal
SMOTE	<i>Sythetic Minority Oversampling Technique</i>
SMOTE-NC	<i>Sythetic Minority Oversampling Technique for Nominal and Continuous</i>
CART	<i>Classsification and Regression Tree</i>
GBM	<i>Gradient Boosting Machine</i>
TP	<i>True Positive</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>
TN	<i>True Negative</i>
LIME	<i>Local Interpretable Model-agnostic Explanations</i>
ICD-10	<i>The International Classification of Diseases, Tenth Revision</i>