

**PENERAPAN ALGORITMA LIGHTGBM DAN ANALISIS
SHAP UNTUK PREDIKSI RISIKO PENYAKIT STROKE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

WISNALA ARYAWIDUTA

22.11.5168

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

PENERAPAN ALGORITMA LIGHTGBM DAN ANALISIS SHAP UNTUK PREDIKSI RISIKO PENYAKIT STROKE

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi S1 Informatika



disusun oleh

WISNALA ARYAWIDUTA

22.11.5168

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**PENERAPAN ALGORITMA LIGHTGBM DAN ANALISIS SHAP UNTUK
PREDIKSI RISIKO PENYAKIT STROKE**

yang disusun dan diajukan oleh

Wisnala Aryawiduta

22.11.5168

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 2 Maret 2026

Dosen Pembimbing,



Dr. Emigawaty, S. Kom., M. Kom.

NIK. 190302226

HALAMAN PENGESAHAN
SKRIPSI
PENERAPAN ALGORITMA LIGHTGBM DAN ANALISIS SHAP UNTUK
PREDIKSI RISIKO PENYAKIT STROKE

yang disusun dan diajukan oleh

Wisnala Aryawiduta

22.11.5168

Telah dipertahankan di depan Dewan Penguji
pada tanggal 2 Maret 2026

Susunan Dewan Penguji

Nama Penguji

Dr. Sri Ngudi Wahyuni, S.T., M.Kom
NIK. 190302060

Ninik Tri Hartanti, S.Kom., M.Kom
NIK. 190302330

Dr. Emigawaty, S.Kom., M.Kom
NIK. 190302226

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 2 Maret 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Wisnala Aryawiduta
NIM : 22.11.5168

Menyatakan bahwa Skripsi dengan judul berikut:

Penerapan Algoritma LightGBM dan Analisis Shap Untuk Prediksi Risiko Penyakit Stroke

Dosen Pembimbing : Dr. Emigawaty, S. Kom., M. Kom.

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 penelitian **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 ketidakbenaran 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, 2 Maret 2026

Yang Menyatakan,



The image shows a blue official stamp of AMIKOM Yogyakarta. The stamp contains the text 'AMIKOM YOGYAKARTA', 'STAMP', and a unique identifier 'GE288ANX145593528'. A handwritten signature in black ink is written over the stamp.

Wisnala Aryawiduta

HALAMAN PERSEMBAHAN

Dengan rasa Syukur kepada Allah SWT atas segala Rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “PENERAPAN ALGORITMA LIGHTGBM DAN ANALISIS SHAP UNTUK PREDIKSI RISIKO PENYAKIT STROKE”. Penyusunan skripsi ini tidak terlepas dari dukungan, doa, bantuan, serta motivasi yang berbagai pihak yang telah memberikan kontribusi baik secara langsung maupun tidak langsung selama proses penyusunan skripsi ini.

Sebagai ungkapan rasa syukur dan terima kasih yang mendalam, penulis mempersembahkan karya sederhana ini kepada:

1. Allah SWT, yang senantiasa memberikan rahmat, kesehatan, kemudahan, serta kekuatan kepada penulis dalam menghadapi setiap proses dan tantangan selama penyusunan skripsi ini.
2. Kedua orang tua tercinta, yang senantiasa memberikan motivasi, semangat, dukungan, kasih sayang, doa yang tiada henti, dukungan moral maupun material, serta pengorbanan yang tidak ternilai sehingga penulis dapat menyelesaikan pendidikan hingga tahap ini.
3. Dosen pembimbing skripsi, yang dengan penuh kesabaran telah memberikan bimbingan, arahan, ilmu, serta motivasi kepada penulis dalam menyelesaikan penelitian ini.
4. Rekan-rekan mahasiswa dan sahabat seperjuangan, yang telah memberikan dukungan, bantuan, kebersamaan, serta semangat selama proses perkuliahan dan penyusunan skripsi.

KATA PENGANTAR

Puji syukur penulis panjatkan atas kehadiran Allah SWT atas segala Rahmat, hidayah, serta karunia-Nya, sehingga skripsi yang berjudul “PENERAPAN ALGORITMA LIGHTGBM DAN ANALISIS SHAP UNTUK PREDIKSI RISIKO PENYAKIT STROKE” dapat diselesaikan dengan. Skripsi ini disusun guna memenuhi salah satu syarat untuk mencapai derajat Sarjana pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Penulis menyadari bahwa penyelesaian skripsi ini, penulis memperoleh banyak bantuan baik secara langsung maupun tidak langsung. Oleh karena itu penulis menyampaikan ucapan terima kasih yang tulus kepada:

1. Prof. Dr. Kusri, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
2. Dr. Emigawaty, S. Kom., M. Kom., selaku Dosen Pembimbing Skripsi yang telah meluangkan waktu, memberikan bimbingan, arahan, serta motivasi yang sangat berharga kepada penulis selama penyusunan karya tulis ini.
3. Tim Dosen Penguji, yang telah memberikan kritik, saran, dan masukan yang membangun guna penyempurnaan skripsi ini.
4. Orang tua dan keluarga tercinta yang senantiasa memberikan doa dan kasih sayang yang tulus sehingga penulis dapat menyelesaikan skripsi ini.
5. Rekan-rekan mahasiswa Program Studi S1 Informatika khususnya kelas Informatika 10 angkatan 22 serta seluruh pihak yang tidak dapat penulis sebutkan satu per satu, yang telah membantu kelancaran penyelesaian penelitian ini.

Penulis berharap agar skripsi ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan, khususnya di bidang informatika dan kesehatan.

Yogyakarta, 8 Februari 2026

Penulis

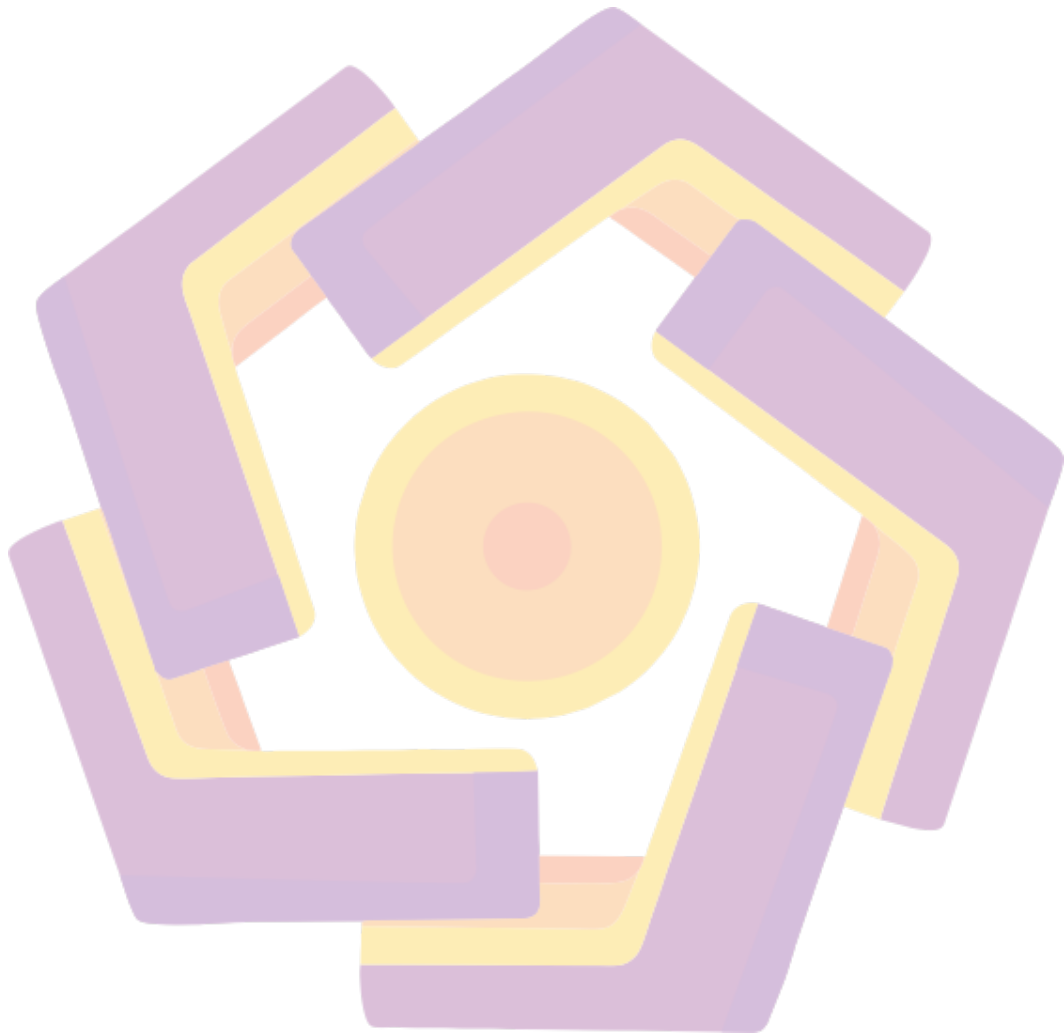
DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI.....	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Studi Literatur	5
2.2 Dasar Teori.....	10
2.2.1 Stroke – Pengertian dan Faktor Risiko.....	10
2.2.2 <i>Machine Learning</i> dan Prediksi Risiko Kesehatan	11
2.2.3 Algoritma LightGBM (Light Gradient Boosting Machine).....	12
2.2.4 SHAP (SHapley Additive Explanations)	13
2.2.5 Evaluasi Model.....	14
BAB III METODE PENELITIAN.....	17

3.1	Objek Penelitian	17
3.2	Alur Penelitian	17
3.3	Alat dan Bahan	20
3.3.1	Data Penelitian	20
3.3.2	Alat dan Instrumen Penelitian	21
BAB IV HASIL DAN PEMBAHASAN		23
4.1	Akuisisi Data	23
4.2	Data Understanding	24
4.3	Hasil Exploratory Data Analysis (EDA)	25
4.4	<i>Preprocessing Data</i>	28
4.5	Implementasi Model LightGBM	30
4.6	Evaluasi Model	31
4.7	Analisis SHAP	33
4.8	Implementasi Prototipe	36
4.9	Pengujian Sistem	36
BAB V PENUTUP		38
5.1	Kesimpulan	38
5.2	Saran	38
REFERENSI		40
LAMPIRAN		42

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian.....	8
Tabel 2. 1 Lanjutan	9

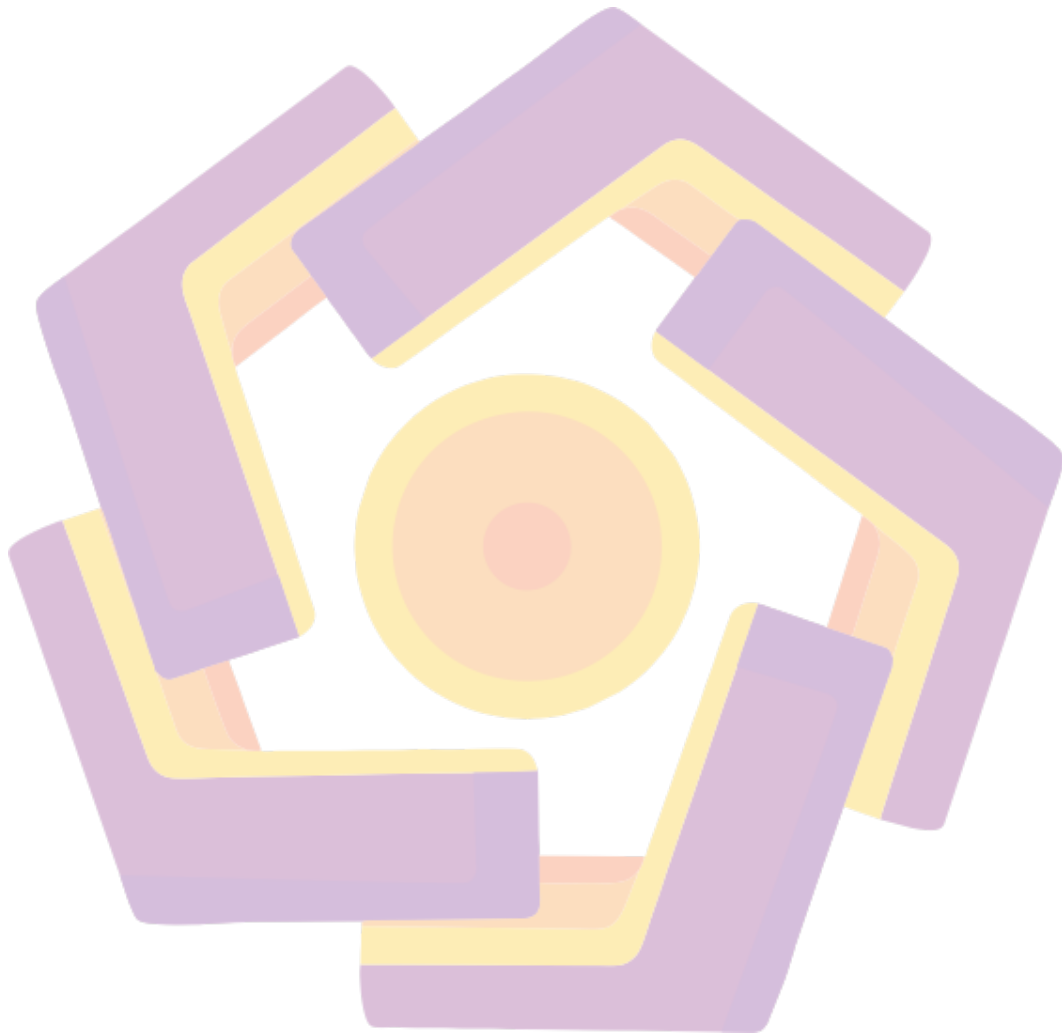


DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	18
Gambar 4. 1 Dataset Stroke	23
Gambar 4. 2 Tipe Data	24
Gambar 4. 3 <i>Missing Value</i>	25
Gambar 4. 4 Distribusi Kelas Stroke	26
Gambar 4. 5 Heatmap Corelation.....	28
Gambar 4. 6 Hasil Label Encoding.....	29
Gambar 4. 7 Hasil Data Splitting	30
Gambar 4. 8 Clasification Report.....	31
Gambar 4. 9 Confusion Matrix	32
Gambar 4. 10 ROC Curve.....	33
Gambar 4. 11 Summary Plot SHAP.....	34
Gambar 4. 12 Prototipe Web.....	36
Gambar 4. 13 Hasil Pengujian Prototipe.....	37

DAFTAR LAMPIRAN

Lampiran 1. Dataset Penelitian	42
Lampiran 2. Repository Program.....	42
Lampiran 3. Implementasi Streamlit.....	42



DAFTAR LAMBANG DAN SINGKATAN

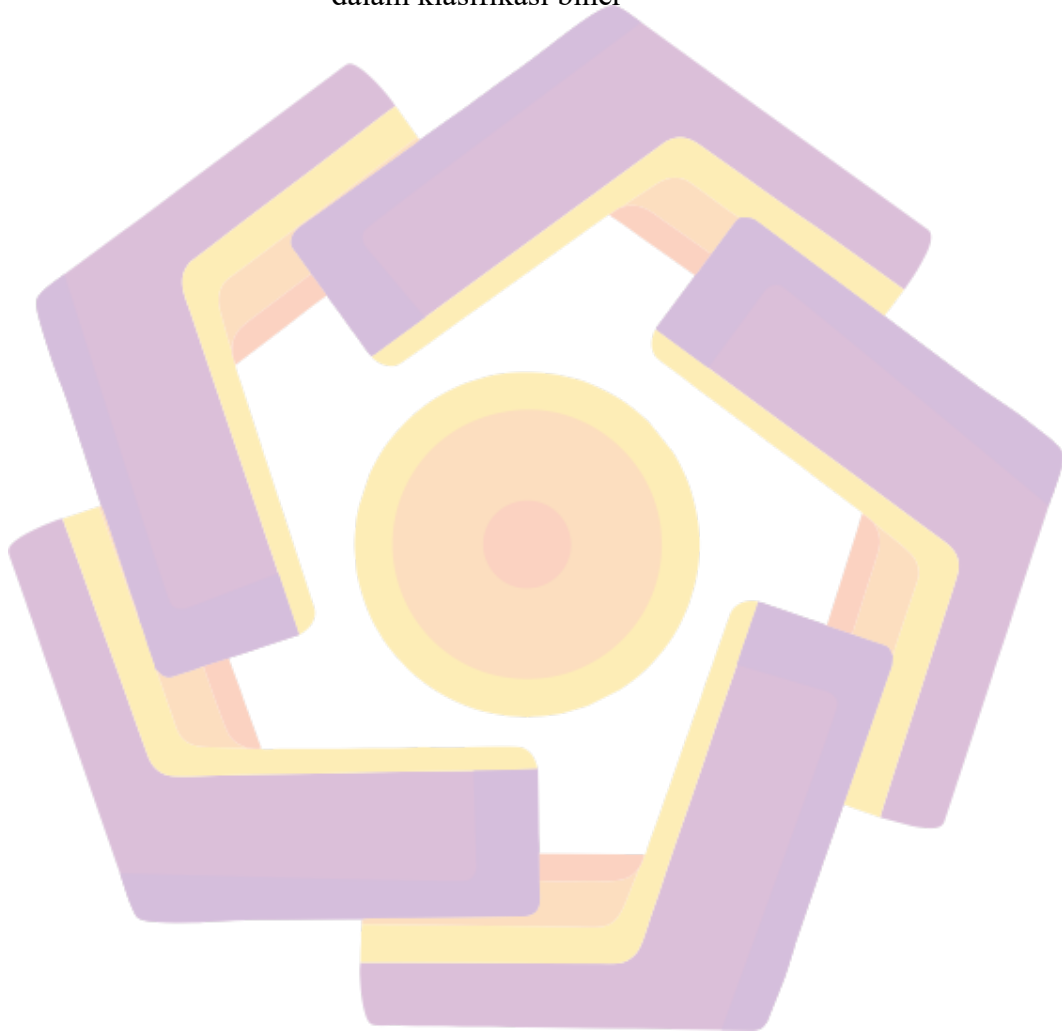


WHO	World Health Organization
LightGBM	Light Gradient Boosting Machine
SHAP	SHapley Additive exPlanations
BMI	<i>Body Mass Index</i> (Indeks Massa Tubuh)
XAI	Explainable Artificial Intelligence
SVM	Support Vector Machines
LIME	Local Interpretable Model-agnostic Explanations
ML	Machine Learning
BPJS	Badan Penyelenggara Jaminan Sosial
XGBoost	eXtreme Gradient Boosting
ANN	Artificial Neural Networks
GOSS	Gradient-based One-Side Sampling
EFB	<i>Exclusive Feature Bundling</i>
EDA	Exploratory Data Analysis
ROC-AUC	<i>Receiver Operating Characteristic Area Under the Curve</i>
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

DAFTAR ISTILAH

<i>Machine Learning</i>	Bidang ilmu komputer yang memberikan kemampuan pada sistem untuk belajar dari data
<i>Decision tree</i>	Metode analisis data dan algoritma <i>machine learning</i> berbentuk struktur pohon hierarkis
<i>Overfitting</i>	Kondisi di mana model terlalu fokus pada data pelatihan sehingga gagal melakukan generalisasi pada data baru
Confusion Matrix	Tabel evaluasi yang membandingkan hasil prediksi model dengan nilai aktual lapangan
<i>Imbalanced Data</i>	Ketidakseimbangan jumlah sampel antar kelas dalam dataset
Label Encoding	Proses konversi variabel kategorikal menjadi numerik
Data Splitting	Pembagian dataset menjadi data latih (<i>training set</i>) dan data uji (<i>testing set</i>)
Leaf-wise	Strategi pertumbuhan pohon pada LightGBM yang memprioritaskan daun dengan tingkat penurunan kerugian terbesar
Hyperparameter	Parameter yang nilainya ditetapkan sebelum proses pelatihan model dimulai untuk mengontrol proses pembelajaran
<i>learning_rate</i>	Parameter yang mengatur seberapa besar langkah yang diambil untuk memperbarui bobot model selama pelatihan
<i>max_depth</i>	Parameter yang mengontrol jarak maksimum antara simpul akar setiap pohon dan simpul daun
Interpretabilitas	Kemampuan untuk memahami dan menjelaskan proses pengambilan keputusan yang mendukung model
Ischemic stroke	jenis stroke yang terjadi ketika akibat adanya gumpalan darah yang menyumbat pembuluh darah di otak
Hemorrhagic stroke	kondisi saat pembuluh darah di otak pecah, menyebabkan perdarahan yang menekan jaringan otak dan menghentikan aliran oksigen

num_leaves	Parameter yang menetapkan jumlah maksimum node per pohon
Depth-wise	Metode pertumbuhan pohon keputusan yang memperluas node berdasarkan tingkat kedalaman secara merata
scale_pos_weight	Paraneter untuk menangani masalah data tidak seimbang dalam klasifikasi biner



INTISARI

Stroke merupakan salah satu penyebab utama kematian dan kecacatan di dunia, dengan prevalensi yang terus meningkat di Indonesia sehingga memerlukan deteksi dini yang akurat guna mencegah dampak kesehatan yang serius. Pendekatan manual konvensional memiliki keterbatasan dalam menangani data kesehatan yang kompleks dan berjumlah besar. Penelitian ini bertujuan untuk menerapkan algoritma *Light Gradient Boosting Machine* (LightGBM) dalam membangun model prediksi risiko stroke serta menggunakan analisis *SHapley Additive exPlanations* (SHAP) untuk memberikan interpretasi terhadap hasil prediksi tersebut. Metode penelitian melibatkan pengolahan Stroke Prediction Dataset yang mencakup 5.110 sampel dengan 12 variabel medis. Tahapan penelitian meliputi *preprocessing* data seperti penanganan nilai hilang pada fitur BMI, *label encoding*, serta pembagian data menjadi set pelatihan dan pengujian dengan rasio 80:20. Model LightGBM dioptimasi menggunakan parameter *scale_pos_weight* untuk menangani ketidakseimbangan kelas agar lebih sensitif dalam mendeteksi kasus positif. Hasil penelitian menunjukkan bahwa model LightGBM mencapai nilai AUC sebesar 0,83 dan *recall* sebesar 82%, yang menandakan kemampuan deteksi yang kuat terhadap pasien berisiko stroke. Analisis SHAP berhasil mengidentifikasi bahwa variabel usia, indeks massa tubuh (BMI), dan kadar glukosa rata-rata merupakan fitur paling berpengaruh terhadap prediksi. Model ini diharapkan dapat menjadi alat bantu bagi tenaga medis dalam mendukung upaya deteksi dini risiko stroke secara lebih transparan dan efisien. Penelitian mendatang disarankan untuk melakukan optimasi hyperparameter lebih lanjut guna meningkatkan presisi pada kelas minoritas.

Kata kunci: Stroke, LightGBM, SHAP, *Machine Learning*, Prediksi.

ABSTRACT

Stroke is a leading cause of death and disability worldwide, with a steadily increasing prevalence in Indonesia, necessitating accurate early detection to prevent serious health impacts. Conventional manual approaches have limitations in handling large and complex health datasets. This study aims to apply the Light Gradient Boosting Machine (LightGBM) algorithm to build a stroke risk prediction model and utilize SHapley Additive exPlanations (SHAP) analysis to provide interpretations for the predictions. The research methodology involves processing the Stroke Prediction Dataset, which includes 5,110 samples with 12 medical variables. The preprocessing stages include handling missing values in the BMI feature, label encoding, and splitting the data into training and testing sets with an 80:20 ratio. The LightGBM model is optimized using the scale_pos_weight parameter to address class imbalance, making it more sensitive in detecting positive cases. The results show that the LightGBM model achieves an AUC value of 0.83 and a recall of 82%, indicating a strong detection capability for patients at risk of stroke. SHAP analysis successfully identifies that age, body mass index (BMI), and average glucose level are the most influential features on the predictions. This model is expected to serve as a tool for medical professionals to support more transparent and efficient early stroke risk detection. Future research is recommended to perform further hyperparameter optimization to improve precision in the minority class.

Keyword: Stroke, LightGBM, SHAP, Machine Learning, Prediction.