



**KOMPARASI ALGORITMA LIGHTGBM, SVM DAN
LOGISTIC REGRESSION DALAM MEMPREDIKSI
PENYAKIT STROKE**

LAPORAN TUGAS AKHIR

BRYANT STEVEN ARITONANG

41521010119

**UNIVERSITAS
MERCU BUANA**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA**

2024



**KOMPARASI ALGORITMA LIGHTGBM, SVM DAN
LOGISTIC REGRESSION DALAM MEMPREDIKSI
PENYAKIT STROKE**

LAPORAN TUGAS AKHIR

BRYANT STEVEN ARITONANG

41521010119

MERCU BUANA

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA**

2024

i

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Bryant Steven Aritonang
NIM : 41521010119
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Komparasi Algoritma LightGBM, SVM, dan Logistic Regression dalam Memprediksi Penyakit Stroke

Menyatakan bahwa Proposal Tugas Akhir ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta, 17 Desember 2024



Bryant Steven Aritonang

UNIVERSITAS
MERCU BUANA

HALAMAN PENGESAHAN





Laporan Skripsi ini diajukan oleh:

Nama : Bryant Steven Aritonang
NIM : 41521010119
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Komparasi Algoritma LightGBM, SVM, dan Logistic Regression dalam Memprediksi Penyakit Stroke

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing : Umniy Salamah, S.T., MMSI
NIDN : 0306098104
Ketua Penguji : Wawan Gunawan, S.Kom., M.T
NIDN : 0424108104
Penguji 1 : Mohamad Yusuf, S.Kom., MCS
NIDN : 0307097606
Penguji 2 : Muhammad Rifqi, S.Kom.,
M.Kom
NIDN : 0301067101

()
()
()
()

UNIVERSITAS
MERCU BUANA

Jakarta, 14 Januari 2025

Mengetahui,

Dekan



Dr. Bambang Jokonowo, S.Si., MTI
NIDN : 0320037002

Ketua Program Studi



Dr. Hadi Santoso, S.Kom., M.Kom
NIDN : 0225067701

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Tugas Akhir ini. Penulisan Tugas Akhir ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Fakultas Ilmu Komputer Universitas Mercu Buana.

Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan Tugas Akhir ini, sangatlah sulit bagi saya untuk menyelesaikan Tugas Akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Orang tua dan Keluarga yang selalu memberikan doa dan dukungan.
2. Bapak Prof. Dr. Andi Adriansyah, M. Eng selaku Rektor Universitas Mercu Buana
3. Bapak Dr. Bambang Jokonowo, S.Si., M.T.I selaku Dekan Fakultas Ilmu Komputer
4. Bapak Dr. Hadi Santoso, S.Kom, M.Kom selaku Ketua Program Studi Teknik Informatika
5. Ibu Umniy Salamah, S.Kom, MMSI selaku Dosen Pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penyusunan proposal skripsi ini.
6. Bapak/Ibu selaku Dosen Penguji Tugas Akhir atas koreksi dan arahan serta masukannya.

Akhir kata, saya berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Tugas Akhir ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 17 Desember 2024



Bryant Steven Aritonang

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Bryant Steven Aritonang
NIM : 41521010119
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Komparasi Algoritma LightGBM, SVM, dan Logistic Regression dalam Memprediksi Penyakit Stroke

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Laporan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 17 Desember 2024

Yang menyatakan,




Bryant Steven Aritonang

UNIVERSIT
MERCU BUANA

ABSTRAK

Nama Mahasiswa : Bryant Steven Aritonang
NIM : 41521010119
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Komparasi Algoritma LightGBM, SVM, dan Logistic Regression dalam Memprediksi Penyakit Stroke.
Pembimbing : Umniy Salamah, S.Kom, MMSI

Stroke merupakan penyakit serius yang dapat menyebabkan kecacatan atau kematian akibat gangguan aliran darah ke otak. Penelitian ini bertujuan untuk membandingkan tiga algoritma machine learning, yaitu LightGBM, Support Vector Machine (SVM), dan Logistic Regression dalam memprediksi risiko stroke. Dataset yang digunakan berjumlah 5110 baris dengan 12 atribut, yang mencakup informasi demografi dan riwayat kesehatan pasien. Proses penelitian dimulai dengan preprocessing data, diikuti dengan pembagian data menjadi data latih dan data uji. Selanjutnya, model dilatih menggunakan ketiga algoritma dan dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil analisis menunjukkan bahwa Logistic Regression memiliki performa terbaik secara keseluruhan dengan keseimbangan antara deteksi kasus stroke dan mengidentifikasi individu yang sehat. SVM menunjukkan hasil yang stabil dengan keseimbangan antara recall dan precision, sementara LightGBM, meskipun memiliki akurasi tinggi, kurang efektif dalam mendeteksi kasus stroke. Penelitian ini menyimpulkan bahwa Logistic Regression adalah model yang paling sesuai untuk memprediksi risiko stroke, meskipun SVM dapat menjadi alternatif yang baik.

Kata Kunci: LightGBM, SVM, Logistic Regression

ABSTRACT

Name : Bryant Steven Aritonang
NIM : 41521010119
Study Program : Informatics Engineer
Thesis Title : Comparison of LightGBM, SVM, and Logistic Regression Algorithms for Stroke Prediction.
Counsellor : Umniy Salamah, S.Kom, MMSI

Stroke is a serious condition that can lead to disability or death due to disrupted blood flow to the brain. This study aims to compare three machine learning algorithms: LightGBM, Support Vector Machine (SVM), and Logistic Regression, in predicting the risk of stroke. The dataset used contains 5110 rows with 12 attributes, including demographic information and health history. The research process began with data preprocessing, followed by splitting the data into training and testing sets. Models were then trained using the three algorithms and evaluated using accuracy, precision, recall, and F1-score metrics. The analysis results indicate that Logistic Regression performed the best overall, providing a balance between detecting stroke cases and identifying healthy individuals. SVM showed stable results with a balance between recall and precision, while LightGBM, despite high accuracy, was less effective in detecting stroke cases. The study concludes that Logistic Regression is the most suitable model for predicting stroke risk, though SVM can be a good alternative.

Keywords: *LightGBM, SVM, Logistic Regression*

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR	iv
HALAMAN PERNYATAAN PERSETUJUAN	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah.....	3
BAB II TINJAUAN PUSTAKA	4
2.1 Penelitian Terdahulu.....	4
2.2 Teori Pendukung	12
2.2.1 LightGBM.....	12
2.2.2 SVM	14
2.2.3 Logistic Regression.....	15
2.2.4 Penyakit Stroke	16

2.2.5	Python	17
2.2.6	Confusion Matrix	18
BAB III METODE PENELITIAN		19
3.1	Jenis Penelitian.....	19
3.2	Tahapan Penelitian	20
BAB IV HASIL DAN PEMBAHASAN		24
4.1	Dataset.....	24
4.2	Eksplorasi Data	25
4.3	Preprocessing Data.....	27
4.4	Split Data.....	34
4.5	Pelatihan Model	36
4.6	Evaluation Model.....	40
BAB V KESIMPULAN DAN SARAN		42
5.1	Kesimpulan.....	42
5.2	Saran.....	44
DAFTAR PUSTAKA		45
LAMPIRAN.....		48



DAFTAR GAMBAR

Gambar 2. 1 Level-wise Tree Growth.....	13
Gambar 2. 2 Leaf-wise Tree Growth	13
Gambar 2. 3 Algoritma Support Vector Machine.....	15
Gambar 2. 4 Kurva Logistic Regression	16
Gambar 3. 1 Tahapan Penelitian	20
Gambar 4. 1 Dataframe	25
Gambar 4. 2 Informasi Atribut.....	26
Gambar 4. 3 Informasi nilai statistik dataframe.....	26
Gambar 4. 4 Missing Value	28
Gambar 4. 5 Handling Missing Value.....	28
Gambar 4. 6 Distribusi Penyakit Stroke.....	29
Gambar 4. 7 Distribusi Fitur Numerik	29
Gambar 4. 8 Distribusi Stroke berdasarkan Fitur Numerik	30
Gambar 4. 9 Plot hubungan Usia dengan Resiko Stroke	31
Gambar 4. 10 Distribusi Fitur Kategorikal untuk mendeteksi Penyakit Stroke....	32
Gambar 4. 11 Output Mean F1 Scores.....	37
Gambar 4. 12 Komparasi Model setelah dilakukan Tuning	39
Gambar 4. 13 Confusion Matrix Model.....	40
Gambar 4. 14 Features Importance Logistic Regression	41

DAFTAR TABEL

Tabel 2. 1 Tinjauan Penelitian	4
Tabel 2. 2 Confusion Matrix	18



DAFTAR LAMPIRAN

Lampiran 1 Kartu Asistensi Bimbingan.....	48
Lampiran 2 Halaman Persetujuan	49
Lampiran 3 Form Revisi Dosen Penguji.....	50
Lampiran 4 Sertifikat BNSP	52
Lampiran 5 Cek Turnitin.....	53
Lampiran 6 Surat Pernyataan HAKI.....	54
Lampiran 7 Submit Jurnal.....	56
Lampiran 8 Curriculum Vitae	57
Lampiran 9 Pernyataan Similarity Check.....	58

