



**IMPLEMENTASI OPTUNA DAN SMOTE PADA OPTIMASI
KLASIFIKASI LIGHTGBM UNTUK KLASIFIKASI DIABETES**

SKRIPSI

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

MUHAMMAD HELMAN AZHARI

NIM 1911016210026

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

DESEMBER 2025



**IMPLEMENTASI OPTUNA DAN SMOTE PADA OPTIMASI
KLASIFIKASI LIGHTGBM UNTUK KLASIFIKASI DIABETES**

SKRIPSI

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

MUHAMMAD HELMAN AZHARI

NIM 1911016210026

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

DESEMBER 2025

SKRIPSI
IMPLEMENTASI OPTUNA DAN SMOTE PADA OPTIMASI
KLASIFIKASI LIGHTGBM UNTUK KLASIFIKASI DIABETES

Oleh:

MUHAMMAD HELMAN AZHARI

1911016210026

Telah dipertahankan di depan dosen penguji pada tanggal 13 Januari 2026.

Susunan Dosen Penguji:

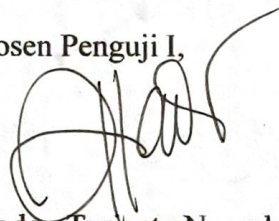
Pembimbing I,



Triando Hamonangan Saragih, S.Kom,
M.Kom.

NIP. 199308242019031012

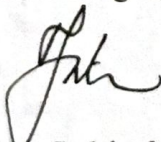
Dosen Penguji I,



Dodon Turiyanto Nugrahadi, S.Kom,
M.Eng.

NIP. 198001122009121002

Pembimbing II,



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

NIP. 198404202008122004

Dosen Penguji II,



Muliadi, S.Kom, M.Sc.

NIP. 197804222010121002

Banjarbaru, 15 Januari 2026

Koordinator Program Studi Ilmu Komputer



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.

Banjarbaru, 20 Desember 2025

Yang Menyatakan,



Muhammad Helman Azhari

NIM. 1911016210026

ABSTRAK

IMPLEMENTASI OPTUNA DAN SMOTE PADA OPTIMASI KLASIFIKASI LIGHTGBM UNTUK KLASIFIKASI DIABETES (Oleh :Muhammad Helman Azhari; Pembimbing: Triando Hamonangan Saragih, S.Kom., M.Kom. dan Fatma Indriani, S.T., M.I.T., Ph.D.; 2025; 60 halaman)

Ketidakseimbangan data merupakan tantangan utama dalam pengembangan model klasifikasi medis, termasuk dalam deteksi dini diabetes. Ketidakseimbangan ini sering menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas dan gagal mengidentifikasi pasien yang benar-benar berisiko. Penelitian ini bertujuan untuk mengoptimalkan kinerja algoritma *Light Gradient Boosting Machine* (LightGBM) dalam klasifikasi diabetes melalui penerapan dua pendekatan, yaitu *Synthetic Minority Oversampling Technique* (SMOTE) untuk penyeimbangan data dan Optuna sebagai metode *hyperparameter tuning*. Dataset yang digunakan dalam penelitian ini adalah *Behavioral Risk Factor Surveillance System* (BRFSS) 2015, yang terdiri dari lebih dari 250.000 sampel dengan proporsi kelas yang tidak seimbang antara responden diabetes dan non-diabetes. Evaluasi dilakukan menggunakan *10-Fold Cross-Validation* dengan berbagai metrik, seperti akurasi, *precision*, *recall*, *F1-score* (*macro* dan *weighted*), serta *Area Under Curve* (AUC). Hasil penelitian menunjukkan kombinasi SMOTE dan Optuna dapat memberikan peningkatan signifikan pada *recall* dan *F1-score*. Hasil evaluasi dari model ini adalah *accuracy* 83,98%, *precision macro* 66,71%, *precision weighted* 84,07%, *recall macro* 66,89%, *recall weighted* 83,98%, *F1-Score macro* 66,79%, *F1-Score weighted* 84,03%, dan AUC 81,80. Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan *hybrid* antara teknik penyeimbangan data dan optimasi *hyperparameter* merupakan strategi efektif dalam menangani permasalahan ketidakseimbangan dataset medis serta meningkatkan performa model klasifikasi diabetes.

Kata kunci: LightGBM, SMOTE, Optuna, *Data Imbalance*, Klasifikasi, Diabetes, BRFSS

ABSTRACT

IMPLEMENTATION OF OPTUNA AND SMOTE FOR OPTIMIZING LIGHTGBM CLASSIFICATION IN DIABETES CLASSIFICATION (By: Muhammad Helman Azhari; Advisor: Triando Hamonangan Saragih, S.Kom., M.Kom. and Fatma Indriani, S.T., M.I.T., Ph.D.; 2025; 60 pages)

Data imbalance is a major challenge in the development of medical classification models, including the early detection of diabetes. This imbalance often causes classification models to be biased toward the majority class and fail to identify patients who are truly at risk. This study aims to optimize the performance of the Light Gradient Boosting Machine (LightGBM) algorithm for diabetes classification by applying two approaches: the Synthetic Minority Oversampling Technique (SMOTE) for data balancing and Optuna for hyperparameter tuning. The dataset used in this study is the Behavioral Risk Factor Surveillance System (BRFSS) 2015, which consists of more than 250,000 samples with an imbalanced class distribution between diabetic and non-diabetic respondents. The evaluation was conducted using 10-fold cross-validation with various metrics, including accuracy, precision, recall, F1-score (macro and weighted), and the Area Under the Curve (AUC). The results show that the combination of SMOTE and Optuna provides a significant improvement in recall and F1-score. The evaluation results of this model are an accuracy of 83.98%, macro precision of 66.71%, weighted precision of 84.07%, macro recall of 66.89%, weighted recall of 83.98%, macro F1-score of 66.79%, weighted F1-score of 84.03%, and an AUC of 81.80. Overall, this study demonstrates that a hybrid approach combining data balancing techniques and hyperparameter optimization is an effective strategy for addressing imbalanced medical datasets and improving the performance of diabetes classification models.

Keywords: LightGBM, SMOTE, Optuna, , Classification, Diabetes, BRFSS

PRAKATA

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul "Implementasi Optuna Dan SMOTE Pada Optimasi Klasifikasi LightGBM Untuk Klasifikasi Diabetes" untuk memenuhi syarat dalam menyelesaikan pendidikan program S-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan shalawat serta salam kehadiran Rasulullah Muhammad SAW beserta para sahabat, kerabat, dan pengikut beliau hingga akhir zaman.

Pada lembar ini penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, Adapun yang dimaksud adalah sebagai berikut:

1. Kedua orang tua dan keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini.
2. Bapak Triando Hamonangan Saragih, S.Kom, M.Kom selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Ibu Fatma Indriani, S.T, M.I.T, Ph.D selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Ibu Dwi Kartini, S.Kom., M.Kom selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2019 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.

7. 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 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 mendapat keridhaan Allah SWT.

Banjarbaru, 16 Desember 2025

Muhammad Helman Azhari

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
ABSTRAK.....	iv
<i>ABSTRACT</i>	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
BAB I PENDAHULUAN.....	13
1.1 Latar Belakang.....	13
1.2 Rumusan Masalah.....	15
1.3 Batasan Masalah.....	16
1.4 Tujuan Penelitian.....	16
1.5 Manfaat Penelitian.....	16
BAB II TINJAUAN PUSTAKA.....	18
2.1 Kajian Terdahulu.....	18
2.2 Keaslian Penelitian.....	21
2.3 Landasan Teori.....	26
2.3.1 <i>Diabetes</i>	26
2.3.2 <i>Data Mining</i>	27
2.3.3 <i>Preprocessing Data</i>	27
2.3.4 <i>Light Gradient-Boosting Machine</i>	28
2.3.5 <i>Hyperparameter Tuning</i>	30
2.3.6 <i>Optuna</i>	31

2.3.7	<i>Synthetic Minority Over-sampling Technique</i>	34
2.3.8	K-Fold Cross-Validation.....	36
2.3.9	<i>Area Under the Curve</i>	37
2.3.10	<i>Evaluasi</i>	37
BAB III METODE PENELITIAN.....		41
3.1	Alat Penelitian	41
3.2	Bahan Penelitian	41
3.3	Variabel Penelitian.....	42
3.4	Alur Penelitian.....	42
BAB IV HASIL DAN PEMBAHASAN		45
4.1	Hasil.....	45
4.1.1	<i>Data collection</i>	45
4.1.2	<i>Preprocessing Data</i>	47
4.1.3	<i>10-Fold Cross Validation</i>	47
4.1.4	SMOTE	48
4.1.5	<i>Hyperparameter Tuning Optuna</i>	48
4.1.6	Klasifikasi	50
4.1.7	Evaluasi.....	50
4.2	Pembahasan	53
BAB V PENUTUP.....		57
5.1	Kesimpulan.....	57
5.2	Saran	57
DAFTAR PUSTAKA		58
LAMPIRAN.....		63

DAFTAR TABEL

Tabel	Halaman
Tabel 1. Keaslian Penelitian.....	21
Tabel 2. Perancangan Penelitian	25
Tabel 3. Evaluasi nilai AUC	37
Tabel 4. Confusion Matrix	38
Tabel 5. Deskripsi Atribut Data	45
Tabel 6. Proporsi Kelas pada Dataset	46
Tabel 7. Contoh Dataset Diabetes Health Indicators	47
Tabel 8. Hyperparameter yang digunakan pada LightGBM.....	49
Tabel 9. Hasil Hyperparameter terbaik menggunakan Optuna pada Model LightGBM + Optuna.....	49
Tabel 10. Hasil Hyperparameter terbaik menggunakan Optuna pada Model LightGBM + SMOTE + Optuna	49
Tabel 11. Hasil Evaluasi LightGBM.....	51
Tabel 12. Hasil uji model LightGBM + SMOTE + Optuna	51
Tabel 13. Perbandingan Metode dan Performa dengan Penelitian Sebelumnya ..	55

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Akumulasi kasus Diabetes Melitus di Dunia	27
Gambar 2. Depth-based tree growth pada algoritma LightGBM.....	30
Gambar 3. Ilustrasi sebaran data pada SMOTE	36
Gambar 4. Ilustrasi K-Fold Cross Validation	36
Gambar 5. Alur Penelitian.....	42
Gambar 6. Distribusi kelas diabetes sebelum dan setelah implementasi SMOTE	48
Gambar 7. Confusion Matrix LightGBM+SMOTE+Optuna.....	53
Gambar 8. Perbandingan Hasil Evaluasi.....	54

DAFTAR LAMPIRAN

Lampiran

Lampiran 1. Source Code Import Library.....	64
Lampiran 2. Source Code Import Dataset.....	64
Lampiran 3. Source Code Checking Missing Value.....	64
Lampiran 4. Source Code Klasifikasi Model LightGBM	65
Lampiran 5. Source Code Klasifikasi Model LightGBM + SMOTE	67
Lampiran 6. Source Code Klasifikasi Model LightGBM + Optuna	69
Lampiran 7. Source Code Klasifikasi Model LightGBM + SMOTE + Optuna ...	73
Lampiran 8. Source Code Hasil Evaluasi	77