PERSEMBAHAN

Dengan mengucap puji syukur kepada Allah S.W.T, Tugas Akhir ini kupersembahkan untuk:

1. Orang tuaku, terima kasih atas dukungan dan semangat yang telah di berikan.
2. Istri dan keempat anakku yang selalu memberiku semangat dalam proses mendapatkan gelar sarjana.

UNIVERSITAS
NUSA MANDIRI

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama	: Thaufik Darma Jonhar
NIM	: 12250068
Program Studi	: Informatika
Fakultas	: Teknologi Informasi
Perguruan Tinggi	: Universitas Nusa Mandiri

Dengan ini menyatakan bahwa Tugas Akhir yang telah saya buat dengan judul: **"Perbandingan Kinerja Algoritma Machine Learning Menggunakan SMOTE Untuk Prediksi Diabetes"**, adalah asli (orsinil) atau tidak plagiat (menjiplak) dan belum pernah diterbitkan/dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya tanpa ada paksanaan dari pihak manapun juga. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa Tugas Akhir yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari **Universitas Nusa Mandiri** dicabut/dibatalkan.

Dibuat di : Jakarta
Pada tanggal : 03 Agustus 2026
Yang menyatakan,



Thaufik Darma Jonhar

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama	: Thaufik Darma Jonhar
NIM	: 12250068
Program Studi	: Informatika
Fakultas	: Teknologi Informasi
Perguruan Tinggi	: Universitas Nusa Mandiri

Dengan ini menyetujui untuk memberikan ijin kepada pihak **Universitas Nusa Mandiri**, Hak Bebas Royalti Non-Eksklusif *(Non-exclusive Royalti-Free Right)* atas karya ilmiah kami yang berjudul: **"Perbandingan Kinerja Algoritma Machine Learning Menggunakan SMOTE Untuk Prediksi Risiko Diabetes"**, beserta perangkat yang diperlukan (apabila ada).

Dengan **Hak Bebas Royalti Non-Eksklusif** ini pihak **Universitas Nusa Mandiri** berhak menyimpan, mengalih-media atau *format*-kan, mengelolaannya dalam pangkalan data *(database)*, mendistribusikannya dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari kami selama tetap mencantumkan nama kami sebagai penulis/pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Universitas Nusa Mandiri , segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 03 Agustus 2026
Yang menyatakan,



Thaufik Darma Jonhar

PERSETUJUAN DAN PENGESAHAN TUGAS AKHIR

Tugas Akhir ini diajukan oleh:

Nama	:	Thaufik Darma Jonhar
NIM	:	12250068
Program Studi	:	Informatika
Fakultas	:	Teknologi Informasi
Jenjang	:	Strata Satu (S1)
Judul Tugas Akhir	:	Perbandingan Kinerja Algoritma Machine Learning Menggunakan SMOTE untuk Prediksi Diabetes

Telah dipertahankan pada periode 2026-1 dihadapan penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh Sarjana Komputer (S.Kom) pada Program Sarjana Program Studi Informatika Fakultas Teknologi Informasi di Universitas Nusa Mandiri.

Jakarta, 18 Agustus 2026

PEMBIMBING TUGAS AKHIR

Dosen Pembimbing : Dr. Ruhul Amin, M.Kom.

Ditandatangani Digital:
Ruhul Amin
2026-08-19 04:34:14

.....

D E W A N P E N G U J I

Penguji I : Frisma Handayanna, M.Kom

Ditandatangani Digital:
Frisma Handayanna
2026-08-18 11:52:00

.....

Penguji II : Sita Anggraeni , M.Kom.

Ditandatangani Digital:
Sita Anggraeni
2026-08-18 10:42:54

.....

PEDOMAN PENGGUNAAN HAK CIPTA

Tugas Akhir yang berjudul **"Perbandingan Kinerja Algoritma Machine Learning Menggunakan SMOTE Untuk Prediksi Diabetes"** adalah hasil karya tulis asli **Thaufik Darma Jonhar** dan bukan hasil terbitan sehingga peredaran karya tulis hanya berlaku di lingkungan akademik saja, serta memiliki hak cipta. Oleh karena itu, dilarang keras untuk menggandakan baik sebagian maupun seluruhnya karya tulis ini, tanpa seizin penulis.

Referensi kepustakaan diperkenankan untuk dicatat tetapi pengutipan atau peringkasan isi tulisan hanya dapat dilakukan dengan seizin penulis dan disertai ketentuan pengutipan secara ilmiah dengan menyebutkan sumbernya.

Untuk keperluan perizinan pada pemilik dapat menghubungi informasi yang tertera di bawah ini:

Nama : Thaufik Darma Jonhar

Alamat : Jalan Zamrud Utara Raya Blok T14/109B, Tambun, Bekasi

No. Telp : 081316859032

E-mail : thaufikdj@gmail.com

UNIVERSITAS NUSA MANDIRI

KATA PENGANTAR

Alhamdulillah, dengan mengucapkan puji syukur kehadirat Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan tugas ini dengan baik. Dimana Tugas Akhir ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul Tugas Akhir, yang penulis ambil sebagai berikut, **"PERBANDINGAN KINERJA ALGORITMA MACHINE LEARNING MENGGUNAKAN SMOTE UNTUK PREDIKSI DIABETES".**

Tujuan penulisan Tugas Akhir ini dibuat sebagai salah satu syarat kelulusan Program Sarjana Universitas Nusa Mandiri. Sebagai bahan penulisan diambil berdasarkan hasil penelitian, observasi dan beberapa sumber literatur yang mendukung penulisan ini. Penulis menyadari bahwa tanpa bimbingan dan dorongan dari semua pihak, maka penulisan Tugas Akhir ini tidak akan lancar. Oleh karena itu pada kesempatan ini, izinkanlah penulis menyampaikan ucapan terima kasih kepada:

1. Rektor Universitas Nusa Mandiri.
2. Wakil Rektor I Bidang Akademik Universitas Nusa Mandiri.
3. Dekan Fakultas Teknologi Informasi Universitas Nusa Mandiri.
4. Ketua Program Studi Informatika Universitas Nusa Mandiri.
5. Bapak Dr. Ruhul Amin, M.kom, selaku Dosen Pembimbing Tugas Akhir.
6. Keluarga ku tercinta yang telah memberikan dukungan moral maupun spritual.
7. Teman-teman mahasiswa kelas 12.8C.10.

Serta semua pihak yang terlalu banyak untuk disebut satu persatu sehingga terwujudnya penulisan ini. Penulis menyadari bahwa penulisan Tugas Akhir ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan dimasa yang akan datang.

Akhir kata semoga Tugas Akhir ini dapat berguna bagi penulis khususnya dan bagi para pembaca yang berminta pada umumnya.

Jakarta, 03 Agustus 2026

Penulis



Thaufik Darma Jonhar

UNIVERSITAS
NUSA MANDIRI

ABSTRAK

Thaufik Darma Jonhar (12250068), Perbandingan Kinerja Algoritma Machine Learning Menggunakan SMOTE Untuk Prediksi Diabetes

Diabetes mellitus sebagai penyakit metabolik kronis yang memerlukan prediksi risiko sejak dini untuk mendukung upaya pencegahan. Adapun tujuannya dari penelitian ini guna melakukan perbandingannya kinerja beberapa algoritma *machine learning* dalam memprediksi risiko diabetes dengan digunakannya teknik *Synthetic Minority Over-sampling Technique* (SMOTE) serta mengimplementasikan model terbaik ke dalam aplikasi web sederhana. *Dataset* yang digunakan adalah *Diabetes Health Indicators* dengan variabel target Diabetes_012, yang terdiri dari tiga kelas, yaitu tidak diabetes, prediabetes, dan diabetes. Tahapan penelitian meliputi pemahaman data, prapemrosesan data, seleksi model awal, pembagian data, pemodelan sebelum SMOTE, penerapan SMOTE pada data latih, pemodelan sesudah SMOTE, evaluasi model, dan implementasi aplikasi web. Algoritma yang dibandingkan ini yaitu *Gradient Boosting*, LightGBM, XGBoost, AdaBoost, dan *Logistic Regression*. Evaluasi kinerja model dilaksanakan dengan digunakannya *accuracy*, *precision*, *recall*, maupun F1-score. Temuan ini menegaskan bahwasanya SMOTE memberikan pengaruh yang berbeda pada setiap algoritma. Beberapa model mengalami peningkatan nilai F1-score setelah penerapan SMOTE, meskipun nilai *accuracy* mengalami penurunan. Performa terbaik diperoleh oleh LightGBM sesudah SMOTE dengan nilai *accuracy* 0,847958, *precision* 0,807324, *recall* 0,847958, maupun F1-score 0,820090. Oleh karena itu, LightGBM sesudah SMOTE dipilih sebagai model terbaik dan diimplementasikan ke dalam prototipe aplikasi web prediksi risiko diabetes. Aplikasi ini memberikan informasi prediksi awal dan tidak digunakan sebagai pengganti diagnosis medis.

Kata Kunci : Diabetes, SMOTE, Prediksi

ABSTRACT

Thaufik Darma Jonhar (12250068), Comparison of Machine Learning Algorithm Performance Using SMOTE for Diabetes Prediction

Early identification of diabetes mellitus risk is essential because this chronic metabolic disease requires timely preventive intervention. For that reason, this research evaluates and compares performance of multiple machine learning algorithms in predicting diabetes risk by incorporating Synthetic Minority Over-sampling Technique (SMOTE), followed by deployment of highest-performing model into a simple web application. The dataset used in this study is the Diabetes Health Indicators dataset with target variable Diabetes_012, which consists of three classes: no diabetes, prediabetes, and diabetes. The research process includes data understanding, preprocessing, initial model selection, data splitting, model training before SMOTE, SMOTE implementation on training data, model training after SMOTE, model evaluation, and web application implementation. The algorithms compared in this study are Gradient Boosting, LightGBM, XGBoost, AdaBoost, and Logistic Regression. Model performance was evaluated using accuracy, precision, recall, and F1-score. The experimental results showed SMOTE affects each algorithm differently. Some models experienced an improvement in F1-score after SMOTE, although their accuracy slightly decreased. The best performance was obtained by LightGBM after SMOTE, with an accuracy of 0.847958, precision of 0.807324, recall of 0.847958, and F1-score of 0.820090. Therefore, LightGBM after SMOTE was selected as best model and implemented into a web application prototype for diabetes risk prediction. The application provides early prediction information and is not intended to replace medical diagnosis.

Keywords : *Diabetes, SMOTE, Prediction*

DAFTAR ISI

LEMBAR JUDUL TUGAS AKHIR i
LEMBAR PERSEMBAHAN ii
LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR iii
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH ... iv
LEMBAR PERSETUJUAN DAN PENGESAHAN TUGAS AKHIR v
LEMBAR PEDOMAN PENGGUNAAN HAK CIPTA vi
KATA PENGANTAR vii
ABSTRAK ix
DAFTAR ISI xi
DAFTAR GAMBAR xiv
DAFTAR TABEL xv
DAFTAR LAMPIRAN xvi
BAB I PENDAHULUAN 1
1.1 Latar Belakang Masalah 1
1.2 Identifikasi Masalah 4
1.3 Rumusan Masalah 4
1.4 Tujuan Penelitian 5
1.5 Ruang Lingkup 6
BAB II LANDASAN TEORI 7
2.1 Tinjauan Pustaka 7
2.1.1 Diabetes mellitus dan Regulasi Insulin 8
2.1.2 Machine learning dan Supervised Learning 9
2.1.3 Automated Machine learning 10
2.1.4 Imbalanced Dataset 10
2.1.5 Synthetic Minority Over-sampling Technique (SMOTE) 11
2.1.6 Algoritma Klasifikasi 12
2.1.7 Metrik Evaluasi 16
2.2 Penelitian Terkait 19
2.2.1 Prediksi Diabetes dengan Machine learning 19
2.2.2 Imbalanced Data dengan SMOTE 20
2.2.3 Ensemble dan Boosting 20
2.2.4 Perbedaan Hasil Penelitian 21
2.2.5 Mapping Penelitian Terdahulu 21

2.2.6 Kebaruan Penelitian 24
2.3 Objek Penelitian 25
2.3.1 Dataset Penelitian 25
2.3.2 Karakteristik Dataset 25
BAB III METODOLOGI PENELITIAN 27
3.1 Tahapan Penelitian 27
3.1.1 Pengumpulan Data 29
3.1.2 Pemahaman Data 30
3.1.3 Prapemrosesan Data 31
3.1.4 Seleksi Model Awal 34
3.1.5 Pembagian Data Latih dan Data Uji 36
3.1.6 Pemodelan sebelum SMOTE 37
3.1.7 Penerapan SMOTE 39
3.1.8 Pemodelan Sesudah SMOTE 41
3.1.9 Evaluasi Model 42
3.1.10 Perbandingan Hasil 44
3.1.11 Penentuan Model Terbaik 46
3.1.12 Persiapan Model untuk Aplikasi Web 48
BAB IV HASIL PENELITIAN DAN PEMBAHASAN 50
4.1 Hasil Penelitian 50
4.1.1 Deskripsi Dataset 50
4.1.2 Distribusi Dataset 51
4.1.3 Hasil Seleksi Model Awal 52
4.1.4 Hasil Evaluasi Model Sebelum SMOTE 53
4.1.5 Hasil Penerapan SMOTE 53
4.1.6 Hasil Evaluasi Model Sesudah SMOTE 54
4.1.7 Perbandingan Kinerja Model Sebelum dan Sesudah SMOTE 55
4.1.8 Penentuan Model Terbaik 56
4.1.9 Confusion matrix Model Terbaik 57
4.1.10 Analisis Kontribusi Fitur 57
4.2 Pembahasan 59
4.2.1 Pengaruh SMOTE terhadap Kinerja Model 59
4.2.2 Analisis Model Terbaik 60
4.2.3 Keterbatasan Hasil Penelitian 60
4.2.4 Keterkaitan Hasil dengan Tujuan Penelitian 61
BAB V PENUTUP 63

UNIVERSITAS NUSA MANDIRI

5.1 Kesimpulan 63
5.2 Saran 64
DAFTAR PUSTAKA 66
DAFTAR RIWAYAT HIDUP 68
LEMBAR BIMBINGAN TUGAS AKHIR 69
LAMPIRAN 70

UNIVERSITAS
NUSA MANDIRI

DAFTAR GAMBAR

Halaman

Gambar III.1 Tahapan Penelitian 28
Gambar III.2 Tahapan Prapemrosesan Data 34
Gambar III.3 Pembagian Data Latih dan Data Uji 36
Gambar III.4 Alur Pemodelan Sebelum SMOTE 38
Gambar III.5 Alur Penerapan SMOTE 40
Gambar III.6 Alur Pemodelan Sesudah SMOTE 41
Gambar III.7 Alur Evaluasi Model 44
Gambar III.8 Alur Perbandingan Hasil Model 45
Gambar III.9 Alur Penentuan Model Terbaik 47
Gambar IV.1 Distribusi Kelas Target Diabetes_012 52
Gambar IV.2 Distribusi Data Latih Sebelum dan Sesudah SMOTE 54
Gambar IV.3 Perbandingan F1-score Sebelum dan Sesudah SMOTE 56
Gambar IV.4 Feature Importance pada Model LightGBM 58

UNIVERSITAS
NUSA MANDIRI

DAFTAR TABEL

Halaman

Tabel II.1 Confusion matrix 17
Tabel II.2 Mapping Penelitian Terdahulu 22
Tabel III.1 Definisi Variabel Dataset 30
Tabel III.2 Algoritma Penelitian 35
Tabel III.3 File Pendukung Implementasi Aplikasi Web 48
Tabel IV.1 Deskripsi Umum Dataset 50
Tabel IV.2 Distribusi Kelas Target Diabetes_012 51
Tabel IV.3 Hasil Seleksi Model Awal Menggunakan PyCaret 52
Tabel IV.4 Hasil Evaluasi Model Sebelum SMOTE 53
Tabel IV.5 Distribusi Data Latih Sebelum dan Sesudah SMOTE 53
Tabel IV.6 Hasil Evaluasi Model Sesudah SMOTE 54
Tabel IV.7 Matrix Comparation Sebelum dan Sesudah SMOTE 55
Tabel IV.8 Model Terbaik Berdasarkan Hasil Evaluasi 56
Tabel IV.9 Confusion matrix Model LightGBM Sesudah SMOTE 57
Tabel IV.10 Feature Importance pada Model LightGBM 58

UNIVERSITAS
NUSA MANDIRI

DAFTAR LAMPIRAN

Halaman

Lampiran A.1. Bukti Hasil Pengecekan Plagiarisme Bab I 70
Lampiran A.2. Bukti Hasil Pengecekan Plagiarisme Bab II 71
Lampiran A.3. Bukti Hasil Pengecekan Plagiarisme Bab III 72
Lampiran A.4. Bukti Hasil Pengecekan Plagiarisme Bab IV 73
Lampiran A.5. Bukti Hasil Pengecekan Plagiarisme Bab V 74
Lampiran B.1. Bukti Submit/Publish Artikel Ilmiah 75
Lampiran C.1. Form Dataset 76

UNIVERSITAS
NUSA MANDIRI