

**PENERAPAN METODE ANOVA F CLASSIF DALAM
PERBANDINGAN ALGORITMA RANDOM FOREST
DAN GRADIENT BOOSTING MACHINE UNTUK
PREDIKSI LEVEL OBESITAS**

SKRIPSI SARJANA INFORMATIKA

Oleh
Muammar Rosulli
217064516044



**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL
2024**

**PENERAPAN METODE ANOVA F CLASSIF DALAM
PERBANDINGAN ALGORITMA RANDOM FOREST
DAN GRADIENT BOOSTING MACHINE UNTUK
PREDIKSI LEVEL OBESITAS
SKRIPSI SARJANA**

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer dari Fakultas Teknologi Komunikasi dan Informatika

Oleh
Muammar Rosulli
217064516044



**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL
2024**

LEMBAR PENGESAHAN

HALAMAN PENGESAHAN

TUGAS AKHIR

PENERAPAN METODE ANOVA F CLASSIF DALAM PERBANDINGAN
ALGORITMA RANDOM FOREST DAN GRADIENT BOOSTING
MACHINE UNTUK PREDIKSI LEVEL OBESITAS



(Yunan Fauzi Wijaya, S.Kom., MMSI)

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

PENERAPAN METODE ANOVA F CLASSIF DALAM PERBANDINGAN ALGORITMA RANDOM FOREST DAN GRADIENT BOOSTING MACHINE UNTUK PREDIKSI LEVEL OBESITAS

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.



LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

**PENERAPAN METODE ANOVA F CLASSIF
DALAM PERBANDINGAN ALGORITMA
RANDOM FOREST DAN GRADIENT BOOSTING
MACHINE UNTUK PREDIKSI LEVEL
OBESITAS**

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2024-2025 pada tanggal 25 February Tahun 2025

Dosen Pembimbing



Yunan Fauzi Wijaya,

S.Kom., MMSI

NIDN 0329067604

Ketua Program Studi



Ratih Titi Komala

Sari, ST., MM., MMSI

NIDN 0301038302

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Muammar Rosulli
 NPM : 217064516044
 Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
 Program Studi : Informatika
 Tanggal Sidang : 25 February 2025

JUDUL DALAM BAHASA INDONESIA :

PENERAPAN METODE ANOVA F CLASSIF DALAM PERBANDINGAN ALGORITMA RANDOM FOREST DAN GRADIENT BOOSTING MACHINE UNTUK PREDIKSI LEVEL OBESITAS

JUDUL DALAM BAHASA INGGRIS :

APPLICATION OF THE CLASSIF F ANOVA METHOD IN COMPARISON OF THE RANDOM FOREST AND GRADIENT BOOSTING MACHINE ALGORITHM FOR PREDICTING OBESITY LEVEL

TANDA TANGAN DAN TANGGAL

Pembimbing	Ka. Prodi	Mahasiswa
TGL : 28 February 2025	TGL : 28 February 2025	TGL : 28 February 2025
	 RATIH TITI R.S.	

ABSTRAK

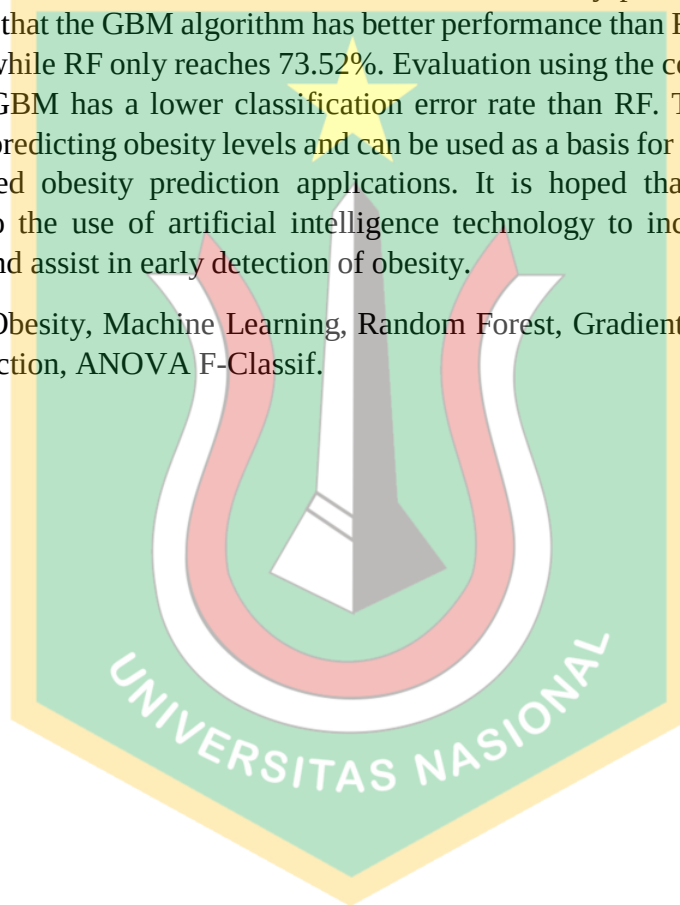
Obesitas merupakan masalah kesehatan global yang terus meningkat dan dapat menyebabkan berbagai penyakit kronis. Dalam penelitian ini, dilakukan perbandingan performa algoritma Random Forest (RF) dan Gradient Boosting Machine (GBM) dalam memprediksi tingkat obesitas berdasarkan dataset yang diperoleh dari Kaggle. Untuk meningkatkan akurasi model, digunakan metode Feature Selection SelectKBest dengan ANOVA F-Classif guna memilih fitur yang paling berpengaruh terhadap prediksi obesitas. Hasil penelitian menunjukkan bahwa algoritma GBM memiliki performa lebih baik dibandingkan RF, dengan akurasi 97,16%, sementara RF hanya mencapai 73,52%. Evaluasi menggunakan confusion matrix juga menunjukkan bahwa GBM memiliki tingkat kesalahan klasifikasi yang lebih rendah dibandingkan RF. Dengan demikian, GBM lebih efektif dalam memprediksi tingkat obesitas dan dapat digunakan sebagai dasar dalam pengembangan aplikasi prediksi obesitas berbasis machine learning. Penelitian ini diharapkan dapat berkontribusi dalam pemanfaatan teknologi kecerdasan buatan untuk meningkatkan kesadaran kesehatan masyarakat dan membantu dalam deteksi dini obesitas.

Kata Kunci: Obesitas, Machine Learning, Random Forest, Gradient Boosting Machine, Feature Selection, ANOVA F-Classif.

ABSTRACT

Obesity is a global health problem that continues to increase and can cause various chronic diseases. In this research, a comparison was made of the performance of the Random Forest (RF) and Gradient Boosting Machine (GBM) algorithms in predicting obesity levels based on datasets obtained from Kaggle. To increase the accuracy of the model, the SelectKBest Feature Selection method with ANOVA F-Classif is used to select the features that have the most influence on obesity prediction. The research results show that the GBM algorithm has better performance than RF, with an accuracy of 97.16%, while RF only reaches 73.52%. Evaluation using the confusion matrix also shows that GBM has a lower classification error rate than RF. Thus, GBM is more effective in predicting obesity levels and can be used as a basis for developing machine learning-based obesity prediction applications. It is hoped that this research can contribute to the use of artificial intelligence technology to increase public health awareness and assist in early detection of obesity.

Keywords: Obesity, Machine Learning, Random Forest, Gradient Boosting Machine, Feature Selection, ANOVA F-Classif.



KATA PENGANTAR

Puji Syukur penulis panjatkan atas kehadiran Allah SWT, yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Penerapan Metode Anova F Classif Dalam Perbandingan Algoritma Random Forest Dan Gradient Boosting Machine Untuk Prediksi Level Obesitas”. Adapun tujuan dari penyusunan skripsi ini ialah untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer di Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional Jakarta.

Dalam penyusunan skripsi ini penulis banyak menerima bantuan, bimbingan serta dukungan dari berbagai pihak, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Maka dalam kesempatan ini penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu dalam penyusunan skripsi ini, terutama kepada yang terhormat:

1. Bapak Dr. Agung Triayudi, S.Kom.,M.Kom Selaku Dekan Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional
2. Ibu Ir. Endah Tri Esti Handayani, MMSI Selaku Wakil Dekan Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional
3. Ibu Ratih Titi Komalasari, ST, MM, MMSI Selaku Ketua Program Studi Informatika Universitas Nasional yang telah mengorbankan waktu, pikiran dan tenaga untuk membimbing serta memberikan saran dalam menyelesaikan skripsi.
4. Para Dosen dan Seluruh Staf akademik Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional yang telah memberikan bekal ilmu yang bermanfaat
5. Para Pimpinan Instansi beserta Mentor yang telah mengizinkan saya untuk melakukan kegiatan penelitian dan memberikan ilmu yang bermanfaat
6. Kedua orangtua tercinta dan saya sayangi yang senantiasa mencurahkan segenap kasih sayang yang tiada henti-hentinya, doa, motivasi, nasehat, serta kesabaran yang begitu besar.

7. Terima kasih kepada Manchester United, klub sepak bola terbaik sepanjang masa dan favorit penulis. Terima kasih telah mengajarkan arti kesetiaan dan kesabaran dalam meraih tujuan, serta pentingnya menghargai setiap proses. Menonton Manchester United, baik di saat kejayaan maupun ketika tim mengalami kekalahan bertubi-tubi, telah mengajarkan penulis arti dari kesetiaan yang tak tergoyahkan. meski di masa sulit, Man United menginspirasi penulis untuk tetap setia dalam menghadapi setiap tantangan kehidupan. Terima kasih telah memberi motivasi untuk terus maju, berusaha, dan memahami bahwa kegagalan serta kehilangan adalah bagian dari proses pembentukan diri. Semangat "Glory-Glory Man United" selalu menjadi dorongan yang menemani penulis selama penulisan skripsi ini.

Penulis mengakui bahwa skripsi ini memiliki kekurangan dalam berbagai aspek, termasuk materi, isi, dan teknik penyajian. Hal ini disebabkan oleh keterbatasan pengetahuan dan kemampuan penulis. Oleh karena itu, penulis sangat menghargai kritik dan saran yang bersifat membangun dari semua pihak, dan berharap agar kontribusi tersebut dapat meningkatkan kualitas skripsi ini ke arah yang lebih baik.

Jakarta, 17 Februari 2025



Muammar Rosulli

DAFTAR ISI

KATA PENGANTAR	i
ABSTRAK	6
LEMBAR PENGESAHAN	2
DAFTAR ISI.....	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	vi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Identifikasi masalah.....	3
1.3 Rumusan Masalah	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	6
1.6 Kontribusi	6
1.7 Batasan Masalah.....	7
BAB II TINJAUAN PUSTAKA.....	9
2.1 Studi Literatur.....	9
2.2 Landasan Teori	18
2.2.1 Obesitas	18
2.2.2 Python	18
2.2.3 Framework Flask.....	19
2.2.4 Machine Learning	19
2.2.5 Algoritma Gradient Boosting Machine	20
2.2.6 Algoritma Random Forest (RF)	21
2.2.7 SelectKBest Anova F value	22
2.2.8 Metode Evaluasi Confusion Matrix	23
BAB III METODOLOGI PENELITIAN.....	26
3.1 Waktu Penelitian	26
3.2 Penentuan Objek Penelitian.....	27

3.3	Fokus Penelitian	27
3.4	Sumber Data	28
3.5	Teknis Pengumpulan Data.....	28
3.6	Dataset	28
3.7	Desain Penelitian.....	30
3.7.1	Tahapan Pendahuluan	31
3.7.2	Tinjauan Tinjauan Pustaka.....	31
3.7.3	Tahapan Pengumpulan Dataset	32
3.7.4	Tahapan Perancangan Model Algoritma GBM dan RF	32
3.7.5	Tahapan Perancangan Aplikasi	33
3.7.6	Tahapan Hasil Aplikasi	34
BAB IV	HASIL DAN PEMBAHASAN	35
4.1	Rancangbangun System Machine Learning	35
4.1.1	Import Dataset	35
4.1.2	Preprocessing	36
4.1.3	SelecKBest Anova F value.....	38
4.1.4	Pelatihan Model	41
4.1.5	Visualisasi Confusion Matrix.....	45
4.2	Perancangan Aplikasi	53
4.2.1	Rancangan Program	53
4.2.2	Tampilan Aplikasi	62
4.3	Pengujian Aplikasi.....	66
BAB V	KESIMPULAN DAN SARAN.....	68
5.1	Kesimpulan.....	68
5.2	Saran	68
	DAFTAR PUSTAKA	70
	LAMPIRAN	73
	Source Code Program.....	73
	Hasil Test Turnitin.....	79
	Bukti Publikasi Jurnal.....	80

DAFTAR GAMBAR

Gambar 3. 1 Desain Penelitian	30
Gambar 4. 1 Source Code Import Dataset.....	35
Gambar 4. 2 Source Code Preprocessing.....	36
Gambar 4. 3 Source Code Process Anova F Value	38
Gambar 4. 4 Source Code Pelatihan Model Algoritma GBM.....	41
Gambar 4. 5 Hasil Akurasi Model Pelatihan Algoritma GBM.....	42
Gambar 4. 6 Source Code Pelatihan Model Algoritma RF	43
Gambar 4. 7 Hasil Akurasi Model Pelatihan Algoritma RF	44
Gambar 4. 8 Source Code Confusion Matrix	46
Gambar 4. 9 Hasil Confusion Matrix Algoritma GBM	47
Gambar 4. 10 Hasil Confusion Matrix Algoritma RF.....	50
Gambar 4. 11 Struktur Program Aplikasi	54
Gambar 4. 12 Source Code Train.py 1.....	55
Gambar 4. 13 Source Code Train.py 2.....	56
Gambar 4. 14 Source Code Train.py 3.....	56
Gambar 4. 15 Source Code app.py 1.....	57
Gambar 4. 16 Source Code app.py 2.....	58
Gambar 4. 17 Source Code app.py 3.....	59
Gambar 4. 18 Source Code app.py 4.....	60
Gambar 4. 19 Source Code app.py 5.....	61
Gambar 4. 20 Tampilan Home.....	62
Gambar 4. 21 Tampilan About.....	63
Gambar 4. 22 Tampilan Form.....	64
Gambar 4. 23 Tampilan Hasil Predict.....	65
Gambar 4. 24 Hasil Pengujian 1.....	66
Gambar 4. 25 Hasil Pengujian 2.....	67

DAFTAR TABEL

Table 2. 1 Studi Literatur	9
Table 2. 2 Confusion Matrix.....	23
Table 4. 1 F Score All Variable	39
Table 4. 2 Variable Selection Anova F Value	40
Table 4. 3 Hasil Akurasi Algoritma Yang Menggunakan metode Anova dan Yang Tidak.	45
Table 4. 4 Hasil Confusion Matrix Algoritma GBM	49
Table 4. 5 Hasil Confusion Matrix Algoritma RF.....	52

