

**ANALISIS FAKTOR YANG MEMPENGARUHI
PRESTASI AKADEMIK MAHASISWA PERGURUAN
TINGGI MENGGUNAKAN RANDOM FOREST DAN
XGBOOST**

SKRIPSI SARJANA SISTEM INFORMASI

Oleh

Muhammad Haikal

227006516127



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL**

2026

**ANALISIS FAKTOR YANG MEMPENGARUHI
PRESTASI AKADEMIK MAHASISWA
PERGURUAN TINGGI MENGGUNAKAN
RANDOM FOREST DAN XGBOOST
SKRIPSI SARJANA**

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar
Sarjana Sistem Informasi dari Fakultas Teknologi Komunikasi

dan Informatika Oleh

Muhammad Haikal

227006516127



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI
DAN INFORMATIKA
UNIVERSITAS NASIONAL**

2026

HALAMAN PENGESAHAN

TUGAS AKHIR

ANALISIS FAKTOR YANG MEMPENGARUHI PRESTASI AKADEMIK
MAHASISWA PERGURUAN TINGGI MENGGUNAKAN RANDOM
FOREST DAN XGBOOST



Muhammad Haikal

227006516127

Dosen Pembimbing 1

A handwritten signature in blue ink, consisting of a large, stylized 'A' followed by a vertical line and a horizontal stroke at the bottom.

Dr. Ir. Andrianingsih, S.Kom., MMSI.

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

ANALISIS FAKTOR YANG MEMPENGARUHI PRESTASI AKADEMIK MAHASISWA PERGURUAN TINGGI MENGGUNAKAN RANDOM FOREST DAN XGBOOST

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi 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.

Jakarta, 2 Maret 2026



Muhammad Haikal

227006516127

LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

ANALISIS FAKTOR YANG MEMPENGARUHI PRESTASI AKADEMIK MAHASISWA PERGURUAN TINGGI MENGGUNAKAN RANDOM FOREST DAN XGBOOST

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2025-2026 pada tanggal 24 Februari Tahun 2026



Dosen Pembimbing 1

Dr. Ir. Andrianingsih, S.Kom., MMSI.

NIDN: 0303097902

Ketua Program Studi

Dr. Ir. Andrianingsih, S.Kom., MMSI.

NIDN: 0303097902

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Muhammad Haikal
NPM : 227006516127
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Sistem Informasi
Tanggal Sidang : 24 Februari 2026

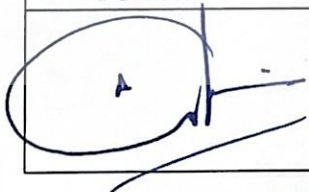


JUDUL DALAM BAHASA INDONESIA :

Analisis Faktor Yang Mempengaruhi Prestasi Akademik Mahasiswa
Perguruan Tinggi Menggunakan Random Forest dan XGBoost

JUDUL DALAM BAHASA INGGRIS :

Analysis of Factors Influencing University Students' Academic
Achievement Using Random Forest and XGBoost

TANDA TANGAN DAN TANGGAL

Pembimbing 1	Ka. Prodi	Mahasiswa
TGL : 2 Maret 2026	TGL : 2 Maret 2026	TGL : 2 Maret 2026
		

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat dan karunia sehingga penulis dapat menyelesaikan skripsi dengan judul “ **ANALISIS FAKTOR YANG MEMPENGARUHI PRESTASI AKADEMIK MAHASISWA PERGURUAN TINGGI MENGGUNAKAN RANDOM FOREST DAN XGBOOST** ” sebagai salah satu syarat kelulusan Program Studi Sarjana Fakultas Teknologi Komunikasi dan Informatika.

Penelitian dan penulisan skripsi ini tidak terlepas dari bantuan berbagai pihak, oleh karena itu penulis menyampaikan banyak terima kasih terutama kepada dosen pembimbing Tugas Akhir, Ibu Dr. Andrianingsih, S.Kom., MMSI. yang telah meluangkan banyak waktu, tenaga, pikiran, arahan, motivasi serta memaklumi segala kekurangan penulis selama penelitian tugas akhir dan penyusunan skripsi.

Penulis juga mengucapkan banyak terima kasih kepada :

1. Prof. Dr. Agung Triayudi, S.Kom., M.Kom., selaku Dekan Fakultas Teknologi Komunikasi dan Informatika.
2. Dr. Ir. Andrianingsih, S.Kom., MMSI., selaku Kepala Program Studi Sistem Informasi dan selaku pembimbing.
3. Seluruh dosen pengajar di Program Studi Sistem Informasi dan dosen di lingkungan Fakultas Teknologi Komunikasi dan Informatika yang telah memberikan ilmu, wawasan, dan pengalaman yang bermanfaat selama masa perkuliahan.
4. Orang tua terkhusus Ibu tercinta yang senantiasa memberikan doa, dukungan, motivasi, dan kasih sayang yang tiada henti kepada penulis.
5. Teman-teman seangkatan dan sehimpuan di Program Studi Sistem Informasi yang telah memberikan dukungan, semangat, serta kebersamaan selama masa perkuliahan dan penyusunan skripsi ini.
6. Sahabat terdekat penulis, yaitu Naufal, Alfad, Ahmad, Ando, Ardan, Fajar, Iqbal dan Pancar, yang senantiasa memberikan dukungan, bantuan, motivasi, serta semangat kepada penulis selama proses perkuliahan hingga penyusunan skripsi ini.

7. Kepada seseorang yang tidak kalah penting kehadirannya, Rizqiyah Wulan Ramadhani. Terima kasih telah menjadi bagian dalam proses perjalanan penulis menyusun skripsi. Berkontribusi baik tenaga, waktu, menemani, mendukung, serta menghibur penulis dalam kesedihan, mendengarkan keluh kesah dan meyakinkan penulis untuk pantang menyerah hingga penyusunan skripsi ini terselesaikan.

Akhir kata, semoga Tuhan Yang Maha Esa membalas kebaikan dan bantuan yang telah diberikan dengan hal yang lebih baik. Penulis mengharapkan kritik dan saran yang bersifat membangun dan semoga skripsi ini dapat memberikan manfaat di bidang Teknologi Informatika.



Jakarta, 20 Februari 2026

Muhammad Haikal

ABSTRAK

Prestasi akademik mahasiswa dipengaruhi oleh faktor akademik, demografis, ekonomi, perilaku belajar, dan dukungan institusi. Penelitian ini bertujuan membangun model klasifikasi menggunakan Random Forest dan XGBoost untuk membedakan mahasiswa unggul ($IPK > 3,5$) dan reguler ($IPK \leq 3,5$), mengevaluasi kinerja model, dan mengidentifikasi faktor dominan melalui feature importance. Penelitian menggunakan pendekatan kuantitatif dengan teknik klasifikasi machine learning pada dataset mahasiswa yang mencakup berbagai variabel akademik dan non-akademik. Evaluasi model menggunakan akurasi, ROC-AUC, precision, recall, dan F1-score. Hasil menunjukkan XGBoost memiliki kinerja lebih baik dengan akurasi 0,75, ROC-AUC 0,77, F1-score macro 0,67, dan F1-score weighted 0,74, dibandingkan Random Forest dengan akurasi 0,73, ROC-AUC 0,804, F1-score macro 0,55, dan F1-score weighted 0,66. Variabel kehadiran_kuliah menjadi faktor paling berpengaruh, diikuti universitas, jenis_kelamin, angkatan, dan bimbingan_dosen. Model diimplementasikan dalam dashboard Streamlit untuk memprediksi klasifikasi dan memberikan rekomendasi perbaikan berdasarkan faktor yang memengaruhi hasil prediksi. Model ini efektif mendukung pemantauan prestasi akademik mahasiswa secara objektif dan berbasis data.

Kata Kunci: Prestasi Akademik, Klasifikasi Prestasi, Feature Importance, Machine Learning, Perguruan Tinggi

ABSTRACT

Student academic performance is influenced by academic, demographic, economic, behavioral, and institutional support factors. This study aims to develop classification models using Random Forest and XGBoost to distinguish high-performing students ($GPA > 3.5$) and regular students ($GPA \leq 3.5$), evaluate model performance, and identify dominant factors using feature importance. A quantitative approach was applied using machine learning classification techniques on student datasets containing academic and non-academic variables. Model evaluation used accuracy, ROC-AUC, precision, recall, and F1-score metrics. The results show that XGBoost achieved better performance with an accuracy of 0.75, ROC-AUC of 0.77, macro F1-score of 0.67, and weighted F1-score of 0.74, compared to Random Forest with an accuracy of 0.73, ROC-AUC of 0.804, macro F1-score of 0.55, and weighted F1-score of 0.66. Class attendance was the most influential factor, followed by university, gender, cohort year, and academic supervision. The model was implemented in a Streamlit dashboard to predict classification and provide improvement recommendations. This model supports objective and data-driven academic performance monitoring.

Keywords: Academic Performance, Performance Classification, Feature Importance, Machine Learning, Higher Education Institutions

DAFTAR ISI

HALAMAN PENGESAHAN.....	i
PERNYATAAN KEASLIAN TUGAS AKHIR.....	ii
LEMBAR PERSETUJUAN REVIEW AKHIR.....	iii
LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU DIREVISI	iv
KATA PENGANTAR.....	v
ABSTRAK.....	vii
ABSTRACT.....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xiii
DAFTAR TABEL.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.2 Research Positioning.....	16
2.2.1 VOSviewer.....	16
2.2.2 Fishbone Diagram.....	17
2.3 Landasan Teori	19
2.3.1 Python	19

2.3.2	Machine Learning	20
2.3.3	Random Forest	21
2.3.4	XGBoost.....	23
2.3.5	Streamlit	25
2.3.6	Prestasi Akademik Mahasiswa.....	26
2.3.7	Faktor yang Mempengaruhi Prestasi Akademik	27
2.4	Kerangka Berpikir	28
2.5	Hipotesis.....	29
2.5.1	Faktor Demografis	29
2.5.2	Faktor Keluarga.....	30
2.5.3	Faktor Pilihan Studi	30
2.5.4	Faktor Ekonomi dan Pekerjaan.....	31
2.5.5	Faktor Kesehatan dan Gaya Hidup	31
2.5.6	Faktor Akademik dan Dosen.....	32
2.5.7	Faktor Fasilitas Kampus.....	32
BAB III METODOLOGI PENELITIAN.....		34
3.1	Jenis Penelitian.....	34
3.2	Lokasi Penelitian.....	34
3.3	Timeline Penelitian	35
3.4	Objek Penelitian	36
3.5	Variabel Penelitian	36
3.6	Sumber Data.....	38
3.7	Pengumpulan Data	38
3.8	Tahapan Penelitian	39
3.8.1	Business Understanding.....	40

3.8.2	Data Understanding.....	41
3.8.3	Data Preparation.....	41
3.8.4	Modeling	43
3.8.5	Evaluasi	48
3.8.6	Deployment.....	51
BAB IV HASIL DAN PEMBAHASAN.....		52
4.1	Business Understanding	52
4.2	Data Understanding.....	53
4.2.1	Struktur Dataset.....	53
4.2.2	Eksplorasi Data (EDA)	54
4.3	Data Preparation.....	57
4.3.1	Data Cleaning.....	57
4.3.2	Penanganan Missing Value	60
4.3.3	Normalisasi dan Pengelompokan Kategorikal	61
4.3.4	Normalisasi dan Semantic Count Encoding Variabel Multilabel..	62
4.3.5	Reverse Negatif Skala Likert	63
4.3.6	Encoding dan Transformasi Variabel Kategorikal dan Ordinal	64
4.3.7	Pelabelan Data.....	65
4.3.8	Analisis Perbedaan Karakteristik Mahasiswa Unggul dan Reguler	67
4.3.9	Split Data.....	67
4.3.10	Data imbalance menggunakan Smote dan Class Weight	68
4.4	Modeling	69
4.4.1	Random Forest	70
4.4.2	XGBoost.....	76

4.5	Evaluasi.....	80
4.5.1	Hasil Evaluasi.....	81
4.5.2	Confusion Matrix	83
4.5.3	Feature Importance.....	84
4.6	Deployment.....	88
BAB V PENUTUP.....		93
5.1	Kesimpulan	93
5.2	Saran.....	94



DAFTAR GAMBAR

Gambar 2. 1 VOSviewer	16
Gambar 2. 2 Fishbone Diagram	18
Gambar 2. 3 Logo Python	20
Gambar 2. 4 Mekanisme Algoritma Random Forest	21
Gambar 2. 5 Mekanisme Algoritma XGBoost.....	24
Gambar 2. 6 Logo Streamlit.....	26
Gambar 2. 7 Alur Kerangka Berpikir.....	28
Gambar 3. 1 Tahapan Penelitian CRISP-DM.....	40
Gambar 4. 1 Responden Duplikat	55
Gambar 4. 2 Distribusi IPK.....	56
Gambar 4. 3 Drop data duplikat.....	58
Gambar 4. 4 Drop Kolom identitas pribadi.....	59
Gambar 4. 5 Distribusi Kolom pendapatan_wali.....	60
Gambar 4. 6 Distribusi Mahasiswa Unggul dan Reguler.....	66
Gambar 4. 7 Perbedaan Mahasiswa Unggul dan Reguler.....	67
Gambar 4. 8 Split data training & testing	68
Gambar 4. 9 Distribusi Sebelum dan Sesudah SMOTE	69
Gambar 4. 10 Tahapan modeling	70
Gambar 4. 11 Alur Modeling Random Forest.....	70
Gambar 4. 12 Hasil evaluasi Random Forest.....	72
Gambar 4. 13 Alur Modeling XGBoost	76
Gambar 4. 14 Hasil evaluasi XGBoost.....	77
Gambar 4. 15 Tahapan Evaluasi.....	80
Gambar 4. 16 Perbandingan akurasi Random Forest & XGBoost	81
Gambar 4. 17 Perbandingan evaluasi kelas Reguler & Unggul.....	82
Gambar 4. 18 Perbandingan AUC Random Forest vs XGBoost	83
Gambar 4. 19 Confusion Matrix Random Forest & XGBoost	84
Gambar 4. 20 Feature Importance Random Forest	85
Gambar 4. 21 Feature Importance XGBoost.....	86

Gambar 4. 22 Tampilan awal ketika membuka streamlit.....	88
Gambar 4. 23 Tampilan ketika dataset sudah upload.....	89
Gambar 4. 24 Halaman Analisis Performa.....	91
Gambar 4. 25 Hasil dari Analisis	92



DAFTAR TABEL

Tabel 2. 1 Tabel Literatur	7
Tabel 3. 1 Timeline Penelitian.....	35
Tabel 3. 2 Variabel Penelitian.....	37
Tabel 4. 1 Tujuan Penelitian.....	52
Tabel 4. 2 Struktur Dataset.....	53
Tabel 4. 3 Format Penulisan IPK	54
Tabel 4. 4 Kolom Missing Value.....	56
Tabel 4. 5 Hasil Penanganan kolom IPK	57
Tabel 4. 6 Penanganan Missing Value.....	60
Tabel 4. 7 Kolom-kolom yang di kelompokkan.....	61
Tabel 4. 8 Hyperparameter Random Forest	71
Tabel 4. 9 Distribusi kelas pada node.....	72
Tabel 4. 10 Hyperparameter XGBoost.....	76
Tabel 4. 11 Struktur pohon pertama model XGBoost.....	78

