

**PREDIKSI RISIKO KETERLAMBATAN KELULUSAN
MAHASISWA S1 PENDIDIKAN TEKNIK BANGUNAN
DI UNJ MENGGUNAKAN MACHINE LEARNING
BERBASIS DATA SIAKAD**



Intelligentia - Dignitas

Disusun sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Program Studi Pendidikan Teknik Bangunan

Oleh:

OCTAVIANUS BONGGO ADHISUKMA

NIM: 1503619054

FAKULTAS TEKNIK

UNIVERSITAS NEGERI JAKARTA

FAKULTAS TEKNIK

PROGRAM STUDI PENDIDIKAN TEKNIK BANGUNAN

TAHUN 2026

LEMBAR PERSETUJUAN SKRIPSI

Nama Mahasiswa : Octavianus Bonggo Adhisukma
NIM : 1503619054
Judul Penelitian : “Prediksi Risiko Keterlambatan Kelulusan Mahasiswa S1 Pendidikan Teknik Bangunan di UNJ Menggunakan Machine Learning Berbasis Data SIAKAD”

ini telah disetujui dan dinyatakan memenuhi syarat untuk diajukan dalam ujian tugas akhir.



Jakarta, 30 Juli 2026
Pembimbing I,

A handwritten signature in blue ink, appearing to read "Murien Nugraheni", with a horizontal line underneath.

Murien Nugraheni, S.T., M.Cs.
NIDN. 0511108701

Pembimbing 2,

A handwritten signature in blue ink, appearing to read "Agphin Ramadhan", with a horizontal line underneath.

Dr. M. Agphin Ramadhan, M.Pd.
NIDN. 0016049004

**LEMBAR PENGESAHAN TUGAS AKHIR
BENTUK : SKRIPSI**

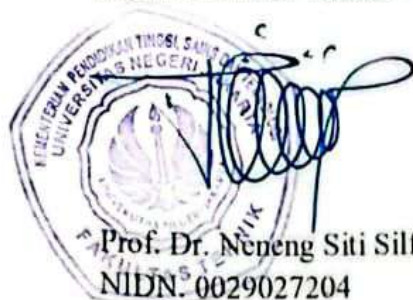
Nama Mahasiswa : Octavianus Bonggo Adhisukma
NIM : 1503619054
Program Studi/Fakultas : Pendidikan Teknik Bangunan / Teknik
Judul Penelitian : "Prediksi Risiko Keterlambatan Kelulusan Mahasiswa
S1 Pendidikan Teknik Bangunan di UNJ Menggunakan
Machine Learning Berbasis Data SIAKAD"

Ini telah dipertahankan dihadapan dewan penguji dan dinyatakan **Lulus**
Ujian Tugas Akhir oleh Tim Penguji pada Tanggal 21 Juli 2026

Tim Penguji,

Ketua Penguji	Dr. Santoso Sri Handoyo, M.T NIDN. 0002126405	
Anggota Penguji (Penguji Ahli 1)	Eka Murtinugraha, M.Pd. NIDN. 0016036706	
Anggota Penguji (Penguji Ahli 2)	Anisah, M.T. NIDN. 0021087505	
Anggota Penguji (Pembimbing 1)	Murien Nugraheni, S.T., M.Cs. NIDN. 0511108701	
Anggota Penguji (Pembimbing 2)	Dr. M. Agphin Ramadhan, M.Pd NIDN. 0016049004	

Mengetahui,
Dekan Fakultas Teknik


Prof. Dr. Neneng Siti Silfi Ambarwati, S.Si., Apt., M.Si.
NIDN. 0029027204

Koordinator Program Studi
Pendidikan Teknik Bangunan


Anisah, M.T
NIDN. 0021087505

LEMBAR PERNYATAAN KEASLIAN KARYA

Nama Mahasiswa : Octavianus Bonggo Adhisukma
NIM : 1503619054
Program Studi : Pendidikan Teknik Bangunan
Fakultas : Teknik
Judul TA : "Prediksi Risiko Keterlambatan Kelulusan Mahasiswa S1 Pendidikan Teknik Bangunan di Unj Menggunakan Machine Learning Berbasis Data SIAKAD"

Menyatakan bahwa Tugas Akhir ini beserta seluruh isinya adalah benar-benar karya saya sendiri. Saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung risiko/sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya tulisilmiah ini.

Jakarta, 30 Juli 2026

Yang menyatakan,



Octavianus Bonggo Adhisukma
1503619054

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa atas segala rahmat, karunia, dan hidayah-Nya, sehingga penulis dapat menyelesaikan penyusunan Skripsi (S1) yang berjudul " PREDIKSI RISIKO KETERLAMBATAN KELULUSAN MAHASISWA S1 PENDIDIKAN TEKNIK BANGUNAN DI UNJ MENGGUNAKAN MACHINE LEARNING BERBASIS DATA SIAKAD". Karya ilmiah ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Pendidikan pada Program Studi Pendidikan Teknik Bangunan, Fakultas Teknik, Universitas Negeri Jakarta.

Penelitian ini bertujuan untuk mengembangkan sebuah model prediksi yang memanfaatkan machine learning untuk menganalisis data SIAKAD guna mengidentifikasi faktor-faktor risiko keterlambatan kelulusan mahasiswa S1 Pendidikan Teknik Bangunan di Universitas Negeri Jakarta. Diharapkan, hasil penelitian ini dapat memberikan kontribusi positif bagi pihak universitas dalam upaya pencegahan dan penanganan isu keterlambatan kelulusan, serta menjadi referensi ilmiah yang bermanfaat bagi pengembangan metode prediksi serupa di bidang akademik maupun industri terkait. Dengan segala kerendahan hati, penulis menyampaikan terima kasih yang tulus kepada:

1. Bapak Prof. Dr. Komarudin, M.Si., selaku Rektor Universitas Negeri Jakarta.
2. Ibu Murien Nugraheni, S.T., M.Cs., selaku Dekan Fakultas Teknik Universitas Negeri Jakarta dan Dosen Pembimbing I, yang senantiasa memberikan arahan, bimbingan, masukan, dan motivasi yang tak ternilai selama proses penyusunan skripsi ini.
3. Ibu Anisah, M.T., selaku Ketua Program Studi Pendidikan Teknik Bangunan Universitas Negeri Jakarta.
4. Bapak Dr. M. Agphin Ramadhan, M.Pd., selaku Dosen Pembimbing II, yang telah meluangkan waktu dan memberikan saran berharga demi kesempurnaan skripsi ini.
5. Seluruh dosen di Program Studi Pendidikan Teknik Bangunan Fakultas Teknik Universitas Negeri Jakarta atas ilmu yang telah diberikan.

6. Kedua orang tua tercinta yang telah memberikan dukungan moral dan material tak terhingga, doa, serta kesabaran yang tiada henti.
7. Teman-teman Pendidikan Teknik Bangunan Universitas Negeri Jakarta serta sahabat-sahabat terdekat yang telah menjadi tempat berbagi cerita, memberikan semangat, bantuan, dan doa selama proses penyusunan skripsi ini berlangsung.
8. Generative AI terutama untuk ChatGPT dalam tata cara penulisan menggunakan frasa yang baik dan juga Claude Code untuk penulisan code yang singkat namun tetap memberikan hasil terbaik

Penulis menyadari bahwa skripsi ini masih memiliki banyak kekurangan dan jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan karya-karya di masa mendatang.

Jakarta, 30 Juli 2026



Octavianus Bonggo Adhisukma

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model prediksi risiko keterlambatan kelulusan mahasiswa S1 Pendidikan Teknik Bangunan di Universitas Negeri Jakarta menggunakan metode *machine learning* berbasis data Sistem Informasi Akademik (SIKAD). Tingginya angka keterlambatan kelulusan, yang mencapai 186 mahasiswa dari total 276 pada data yang diteliti, menggarisbawahi urgensi penanganan isu ini. Metode penelitian ini melibatkan pengumpulan dan pemrosesan data akademik mahasiswa angkatan 2017-2021, mencakup Indeks Prestasi Semester (IPS), jumlah SKS, data mata kuliah, dan biaya kuliah, untuk kemudian dilakukan rekayasa fitur (*feature engineering*). Lima algoritma *machine learning* dievaluasi: *Random Forest*, *Extra Trees*, *Gradient Boosting*, *XGBoost*, dan *CatBoost*. Hasil evaluasi menunjukkan bahwa algoritma *Random Forest* menawarkan performa terbaik dengan keseimbangan antara akurasi, presisi, *recall*, dan kestabilan model, yang dibuktikan melalui *analisis feature importance* di mana IPS semester 2, 3, dan 4 beserta perubahannya (Delta IPS) menjadi faktor prediktif dominan. Model *Random Forest* ini diharapkan dapat berfungsi sebagai *early warning system* bagi program studi untuk mengidentifikasi mahasiswa berisiko terlambat lulus sejak dini, sehingga intervensi akademik yang tepat sasaran dapat dilakukan guna meningkatkan efisiensi dan kualitas lulusan.

Kata Kunci: Prediksi Keterlambatan Kelulusan, *Machine Learning*, SIKAD, *Random Forest*, Indeks Prestasi Semester

ABSTRACT

This study aims to develop and evaluate a machine learning-based model for predicting the risk of delayed graduation among undergraduate students in the Building Engineering Education Program at Universitas Negeri Jakarta using data from the Academic Information System (SIKAD). The high rate of delayed graduation, with 186 out of 276 students in the analyzed dataset experiencing delayed graduation, highlights the urgency of addressing this issue. The research methodology involved collecting and preprocessing academic data from students admitted between 2017 and 2021, including Semester Grade Point Average (GPA), accumulated credit units (SKS), course records, and tuition fee data, followed by a feature engineering process. Five machine learning algorithms were evaluated, namely Random Forest, Extra Trees, Gradient Boosting, XGBoost, and CatBoost. The evaluation results indicate that the Random Forest algorithm achieved the best overall performance by providing a balanced combination of accuracy, precision, recall, and model stability. Feature importance analysis further revealed that the Semester Grade Point Average (GPA) in the second, third, and fourth semesters, along with the changes in GPA (Delta GPA), were the most influential predictive factors. The proposed Random Forest model is expected to serve as an early warning system for the study program to identify students at risk of delayed graduation at an early stage, enabling timely academic interventions to improve graduation efficiency and the overall quality of graduates.

Keywords: Delayed Graduation Prediction, Machine Learning, Academic Information System (SIKAD), Random Forest, Semester Grade Point Average (GPA)

DAFTAR ISI

LEMBAR PERSETUJUAN SKRIPSI.....	ii
LEMBAR PENGESAHAN TUGAS AKHIR.....	iii
LEMBAR PERNYATAAN KEASLIAN KARYA.....	iv
KATA PENGANTAR.....	v
ABSTRAK.....	vii
ABSTRACT.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xvi
BAB I.....	1
1.1. Latar Belakang.....	1
1.2. Identifikasi Masalah.....	4
1.3. Pembatasan Masalah.....	5
1.4. Rumusan Masalah.....	6
1.5. Tujuan Penelitian.....	6
1.6. Manfaat Hasil Penelitian.....	6
BAB II.....	8
2.1. Kerangka Teoritis.....	8
2.2. Konsep Dasar Machine Learning.....	8
2.2.1. Definisi Machine Learning.....	8
2.2.2. Kategori dalam Machine Learning.....	8
2.2.3. Metode Evaluasi Umum.....	11
2.3. Algoritma Machine Learning.....	13
2.3.1. Random Forest.....	13
2.3.2. Extra Trees.....	14
2.3.3. Gradient Boosting.....	15
2.3.4. XGBoost.....	17

2.3.5.	CatBoost	18
2.4.	Prediksi Risiko Keterlambatan Kelulusan	20
2.4.1.	Definisi “Keterlambatan Kelulusan”	20
2.4.2.	Faktor yang Memengaruhi Keterlamabatan Kelulusan.....	21
2.5.	Sistem Informasi Akademik (SIKAD) sebagai Sumber Data Mining	21
2.5.1.	Gambaran Umum SIKAD	21
2.5.2.	Variabel Akademik yang Digunakan.....	22
2.6.	Variable dan Feature Engineering untuk Prediksi.....	22
2.6.1.	Pembuatan Feature Engineering	22
2.6.2.	Feature Importance.....	23
2.7.	Penelitian Terdahulu yang Relevan.....	23
2.8.	Kerangka Berpikir	25
BAB III		27
3.1.	Tempat dan Waktu Penelitian.....	27
3.2.	Alat dan Bahan Penelitian.....	27
3.2.1.	Alat Penelitian.....	27
3.2.2.	Bahan Penelitian.....	27
3.3.	Diagram Alir Penelitian.....	29
3.3.1.	Pengumpulan Data	30
3.3.2.	Seleksi dan Pemahaman Data	30
3.3.3.	Pra-pemrosesan Data.....	31
3.3.4.	Pembagian Data	57
3.3.5.	Pembangunan Model.....	57
3.3.6.	Evaluasi Performa Model.....	63
3.3.7.	Analisis Kestabilan Model.....	65
3.3.8.	Perbandingan Performa dan Stabilitas Model.....	67
3.3.9.	Analisis Feature Importance	69
3.3.10.	Implementasi Model Terbaik ke Aplikasi Streamlit.....	73
3.3.11.	Pengujian Fungsional Aplikasi.....	74

3.3.12.	Analisis Hasil dan Pembahasan	74
3.4.	Teknik dan Prosedur Pengumpulan Data	75
3.4.1.	Teknik Pengumpulan Data	75
3.4.2.	Prosedur Pengumpulan Data	75
3.5.	Teknik Analisis Data	76
BAB IV		79
4.1.	Hasil Penelitian	79
4.1.1.	Pengumpulan Data	79
4.1.2.	Seleksi dan Pemahaman Data	79
4.1.3.	Pra-pemrosesan Data.....	80
4.1.4.	Pembagian Data	83
4.1.5.	Pembangunan Model.....	83
4.1.6.	Evaluasi Performa Model.....	84
4.1.7.	Analisis Kestabilan Model.....	85
4.1.8.	Perbandingan Performa dan Kestabilan Model.....	87
4.1.9.	Analisis Feature Importance	87
4.2.	Implementasi Sistem Prediksi	89
4.2.1.	Arsitektur Sistem.....	89
4.2.2.	Tampilan Sistem.....	90
4.2.3.	Hasil Prediksi	91
4.2.4.	Pedoman Pengguna	92
4.3.	Analisis Hasil dan Pembahasan	93
BAB V		102
5.1.	Kesimpulan	102
5.2.	Saran.....	103
DAFTAR PUSTAKA		105
LAMPIRAN		109

DAFTAR TABEL

Tabel 2. 1 Metode Evaluasi Umum.....	12
Tabel 4. 1 Pembagian Data Latih dan Data Uji.....	83
Tabel 4. 2 Hasil Evaluasi Performa Model Machine Learning.....	85
Tabel 4. 3 Hasil Analisis Kestabilan Model	86
Tabel 4. 4 Distribusi Kelas Variabel Lulus_Terlambat	93
Tabel 4. 5 Distribusi Status Kelulusan Berdasarkan Jenis Kelamin	94
Tabel 4. 6 Distribusi Status Kelulusan Berdasarkan Angkatan	95
Tabel 4. 7 Sepuluh Mata Kuliah dengan Jumlah Nilai D/E Terbanyak pada Mahasiswa Lulus Terlambat.....	96
Tabel 4. 8 Distribusi Kategori IPS Semester 1 Berdasarkan Status Kelulusan.....	97
Tabel 4. 9 Distribusi Kategori IPS Semester 2 Berdasarkan Status Kelulusan.....	98
Tabel 4. 10 Distribusi Kategori IPS Semester 3 Berdasarkan Status Kelulusan.....	98
Tabel 4. 11 Distribusi Kategori IPS Semester 4 Berdasarkan Status Kelulusan.....	98



DAFTAR GAMBAR

Gambar 2. 1 Algoritma Supervised Learning	9
Gambar 2. 2 Algoritma Unsupervised Learning	10
Gambar 2. 3 Algoritma Reinforcement Learning	11
Gambar 2. 4 Arsitektur Random Forest	14
Gambar 2. 5 Arsitektur Extra Trees.....	15
Gambar 2. 6 Arsitektur Gradient Boosting	17
Gambar 2. 7 Arsitektur XGBoost.....	18
Gambar 2. 8 Arsitektur CatBoost.....	19
Gambar 2. 9 Kerangka Berpikir.....	26
Gambar 3. 1 Diagram Alir Penelitian.....	29
Gambar 3. 2 Tabel dataset <i>mahasiswa_aktif_ptb.xlsx</i>	31
Gambar 3. 3 Tabel dataset <i>nilai_matkul_ptb.xlsx</i>	31
Gambar 3. 4 Proses Penggabungan dataset mahasiswa-aktif-ptb.xlsx	32
Gambar 3. 5 Proses Pembersihan dan Pengurutan Data	33
Gambar 3. 6 Proses Penyimpanan Dataset ke data_mahasiswa_aktif.xlsx.....	33
Gambar 3. 7 Proses Penggabungan nilai-matkul-ptb.xlsx	34
Gambar 3. 8 Proses Pembersihan dan Penyusunan Data Nilai Mata Kuliah	35
Gambar 3. 9 Proses Penyimpanan Dataset ke data_nilai_matkul.xlsx	35
Gambar 3. 10 Seleksi Mahasiswa yang Telah Lulus Skripsi	36
Gambar 3. 11 Penggabungan Data Nilai Mata Kuliah dengan Data Mahasiswa Aktif	37
Gambar 3. 12 Seleksi Data Berdasarkan Angkatan Mahasiswa	38
Gambar 3. 13 Imputasi Nilai Biaya Kuliah Menggunakan Modus.....	38
Gambar 3. 14 Penyimpanan Hasil Integrasi Data ke dataset data_mix.xlsx.....	38
Gambar 3. 15 Pembacaan Dataset Penelitian.....	39
Gambar 3. 16 Persiapan Data Sumber	39
Gambar 3. 17 Penentuan Status Keterlambatan Kelulusan.....	41
Gambar 3. 18 Pembentukan Informasi Biaya Kuliah Mahasiswa	42
Gambar 3. 19 Pembentukan Informasi Jenis Kelamin Mahasiswa.....	43

Gambar 3. 20 Seleksi Data Semester 1–4	44
Gambar 3. 21 Pembentukan Identitas Mahasiswa	45
Gambar 3. 22 Ekstraksi Informasi IPS dan SKS Semester 1–4	46
Gambar 3. 23 Ekstraksi Nilai IPK Kumulatif	47
Gambar 3. 24 Pengelompokan Mata Kuliah Berdasarkan Kategori	48
Gambar 3. 25 Perhitungan Jumlah Mata Kuliah Lulus Berdasarkan Kategori	49
Gambar 3. 26 Perhitungan Jumlah Nilai D dan E	50
Gambar 3. 27 Pembentukan Atribut Pembayaran dan Golongan UKT	51
Gambar 3. 28 Penyimpanan Hasil Ekstraksi Fitur	52
Gambar 3. 29 Pembentukan Dataset Akhir	52
Gambar 3. 30 Pembentukan Atribut SKS Maksimum Semester Pertama	53
Gambar 3. 31 Perhitungan Delta IPS dan SKS Maksimum	54
Gambar 3. 32 Penyusunan Kelompok Kolom Dataset	55
Gambar 3. 33 Pengaturan Ulang Urutan Kolom Dataset	56
Gambar 3. 34 Penyimpanan Dataset Penelitian ke dataset_penelitian.xlsx	56
Gambar 3. 35 Pembagian Dataset Menjadi Data Pelatihan dan Data Pengujian	57
Gambar 3. 36 Penentuan Variabel Target dan Variabel Prediktor	58
Gambar 3. 37 Seleksi Fitur yang Tidak Digunakan	58
Gambar 3. 38 Identifikasi Tipe Data Fitur	59
Gambar 3. 39 Penyusunan Pipeline Prapemrosesan Data	60
Gambar 3. 40 Transformasi Data Pelatihan dan Pengujian	61
Gambar 3. 41 Inisialisasi Model Machine Learning	62
Gambar 3. 42 Proses Pelatihan dan Prediksi Model	63
Gambar 3. 43 Perhitungan Metrik Evaluasi Model	64
Gambar 3. 44 Penyajian dan Perankingan Hasil Evaluasi Model	65
Gambar 3. 45 Pengujian Stabilitas Model Menggunakan Repeated Stratified K-Fold Cross Validation	66
Gambar 3. 46 Penyajian Hasil Pengujian Stabilitas Model	67
Gambar 3. 47 Pemilihan Model Terbaik Secara Manual	68
Gambar 3. 48 Pembentukan dan Penyimpanan Pipeline Model	69

Gambar 3. 49 Evaluasi Model Menggunakan Confusion Matrix	70
Gambar 3. 50 Pembentukan Data Feature Importance	71
Gambar 3. 51 Agregasi Nilai Feature Importance	72
Gambar 3. 52 Visualisasi Feature Importance	73
Gambar 4. 1 Hasil Penggabungan Dataset Mahasiswa Aktif dan Nilai Mata Kuliah .	81
Gambar 4. 2 Hasil Penggabungan Dataset Mahasiswa Aktif dan Nilai Mata Kuliah .	81
Gambar 4. 3 Dataset Penelitian Hasil Feature Engineering	82
Gambar 4. 4 Dataset Penelitian Hasil Feature Engineering	82
Gambar 4. 5 Dataset Penelitian Hasil Feature Engineering	82
Gambar 4. 6 Dataset Penelitian Hasil Feature Engineering	82
Gambar 4. 7 Confusion Matrix Model Random Forest	88
Gambar 4. 8 Feature Importance Model Random Forest	89
Gambar 4. 9 Tampilan Utama Sistem Prediksi	90
Gambar 4. 10 Hasil Prediksi Risiko Keterlambatan Kelulusan Menggunakan Model Random Forest	91



DAFTAR LAMPIRAN

Lampiran 1 Surat Tugas	109
Lampiran 2 Lembar Konsultasi Dosen Pembimbing ke 1	110
Lampiran 3 Lembar Konsultasi Dosen Pembimbing ke 1	111
Lampiran 4 Lembar Konsultasi Dosen Pembimbing ke 2	112
Lampiran 5 Lembar Konsultasi Dosen Pembimbing ke 2	113
Lampiran 6 Surat Penelitian.....	114

