



Prediksi Dampak Keaktifan Organisasi Terhadap Keterlambatan Studi Mahasiswa Menggunakan Gradient Boosting

Riski Ananda¹, Dicky Apdillah²

^{1,2} Teknik Informatika, Fakultas Teknik, Universitas Asahan

ra6570225@gmail.com

Abstrak

Penelitian ini bertujuan untuk menganalisis pengaruh keaktifan organisasi terhadap keterlambatan studi mahasiswa serta membangun model prediksi menggunakan algoritma Gradient Boosting. Objek penelitian adalah mahasiswa Program Studi Teknik Informatika Universitas Asahan angkatan 2020–2022 dengan total 151 sampel. Data dikumpulkan melalui Sistem Informasi Akademik (SIKAD) untuk variabel akademik dan kuesioner untuk variabel keaktifan organisasi. Variabel akademik meliputi IPK kumulatif, jumlah SKS, mata kuliah remedial, kehadiran, dan cuti akademik, sedangkan variabel organisasi mencakup jenis organisasi, jabatan, lama keaktifan, frekuensi kegiatan, serta skor keaktifan sebagai hasil rekayasa fitur. Metode yang digunakan adalah Gradient Boosting dengan implementasi XGBoost yang dioptimasi menggunakan Grid Search Cross Validation. Hasil penelitian menunjukkan adanya korelasi positif signifikan antara keaktifan organisasi dan keterlambatan studi ($r = 0,612$; $p < 0,001$). Model terbaik menghasilkan akurasi 90,32%, presisi 90,60%, recall 90,60%, F1-Score 90,57%, dan AUC-ROC 94,87%, serta menunjukkan stabilitas melalui 10-fold cross validation dengan akurasi rata-rata 89,80%. Model ini juga mengungguli algoritma lain seperti Random Forest, KNN, Decision Tree, Logistic Regression, dan Naive Bayes. Analisis feature importance menunjukkan bahwa skor keaktifan organisasi merupakan faktor paling dominan dalam prediksi. Hasil penelitian ini merekomendasikan penerapan sistem peringatan dini untuk mendeteksi mahasiswa berisiko keterlambatan studi secara lebih efektif.

Kata Kunci: Keterlambatan Studi, Keaktifan Organisasi, Machine Learning, Gradient Boosting, Xgboost, Prediksi, Feature Importance, Universitas Asahan

1. Pendahuluan

Pendidikan tinggi merupakan salah satu investasi terpenting dalam kehidupan seorang individu serta dalam menghasilkan sumber daya manusia yang berkualitas dan kompetitif. Salah satu indikator keberhasilan pendidikan diperguruan tinggi yaitu membimbing ketepatan waktu mahasiswa dalam menyelesaikan studi. Namun, pada kenyataannya saat ini masih banyak mahasiswa yang mengalami keterlambatan dalam menyelesaikan studinya tepat waktu dikarenakan adanya pengaruh dari berbagai faktor baik akademik maupun non akademik. Pendidikan tinggi adalah yang menyelenggarakan pendidikan akademik dan dapat menyelenggarakan pendidikan vokasi dalam satu rumpun ilmu pengetahuan dan teknologi tertentu jika memenuhi syarat, pendidikan tinggi dapat dapat menyelenggarakan pendidikan profesi (Septiani et al., 2024).

Salah satu faktor non akademik yang sering dikaitkan dengan keterlambatan studi adalah keaktifan mahasiswa dalam organisasi kemahasiswaan, yang meskipun ikut serta dalam organisasi memberikan manfaat seperti peningkatan *soft skill*, dan menambah pengalaman sosial dan relasi, di sisi lain aktivitas organisasi yang berlebihan diduga dapat mengurangi fokus mahasiswa dalam kegiatan belajar sehingga berpotensi menyebabkan keterlambatan dalam menyelesaikan studi yang harusnya tepat waktu.

Seiring dengan perkembangan teknologi, metode data *mining* dan *machine learning* dapat digunakan untuk memprediksi keterlambatan studi mahasiswa secara lebih efektif. Salah satu metode yang dipilih memiliki performa tinggi dalam prediksi yaitu *Gradient Boosting*, yang merupakan *algoritma ensemble learning* yang mampu meningkatkan akurasi prediksi dengan menggabungkan beberapa model yang lemah menjadi model yang lebih kuat. Dengan menggunakan metode model *Gradient Boosting* ini memprediksi dampak keaktifan organisasi terhadap keterlambatan studi mahasiswa secara lebih akurat dan objektif.

Pada penelitian terdahulu yang dilakukan oleh (Moh. Saiful Hakiki Dkk., 2023), yang berjudul Prediksi Kelulusan Mahasiswa Menggunakan *Algoritma XG Boost*, yang mengatakan bahwa model *XGBoost* ini dapat menjadi alat bantu bagi institusi dalam memantau dan mengintervensi potensi keterlambatan mahasiswa sejak dini.

Penelitian oleh (Alam Dkk., 2024), yang berjudul *Gradient Boosting Machine* pada Klasifikasi Kelulusan Mahasiswa, yang menyimpulkan bahwa penggunaan *Gradient Boosting Machine* mampu meningkatkan prediksi model dalam klasifikasi kelulusan mahasiswa yang dilakukan oleh Prodi Statistika UNTAN.

Penelitian oleh (Yurika Dkk., 2026), yang berjudul Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Perbandingan *Algoritma Random Forest* dan *XGBoost*, yang mengatakan keunggulan dalam menggunakan model

XGBoost berasal dari kemampuannya yang dapat mengoptimalkan melalui sistem *gradient boosting* yang secara iteratif memperbaiki kesalahan dari pohon keputusan sebelumnya.

Pada penelitian sebelumnya telah menunjukkan bahwa metode *Gradient Boosting* akurat dalam memprediksi suatu fenomena yang bersifat pasti. Sehingga penerapan metode tersebut memiliki kemampuan yang baik dalam melakukan prediksi kelulusan mahasiswa. Oleh karena itu, model tersebut terbukti efektif dalam membantu institusi pendidikan untuk memantau serta mengidentifikasi potensi keterlambatan studi sejak dini. Selain itu, metode Gradient Boosting mampu meningkatkan akurasi prediksi melalui proses iteratif dalam memperbaiki kesalahan dari model sebelumnya, algoritma ini juga sangat relevan digunakan dalam penelitian terkait prediksi dampak keaktifan organisasi terhadap keterlambatan studi mahasiswa. Namun, sampai saat ini belum ada penelitian yang menggunakan metode *Gradient Boosting* secara spesifik untuk memprediksi di Universitas Asahan.

2. Metode Penelitian

Dalam penelitian ini, metode yang digunakan meliputi machine learning dan penerapan algoritma gradient boosting untuk menganalisis terjadinya keterlambatan akademik studi mahasiswa dengan pengaruh organisasi dalam akademik mereka. Metode-metode yang dilakukan untuk memperoleh pemahaman teori terkait kenapa mahasiswa bisa mengalami keterlambatan dalam masalah akademik, indikator prestasi akademik dan organisasi serta konsep dan implementasi algoritma machine gradient booster. Algoritma Gradient Boosting digunakan sebagai pendekatan utama dalam proses yang bertujuan membentuk pohon keputusan berdasarkan kesalahan yang dilakukan oleh metode sebelumnya, sehingga dapat diperoleh pola yang menjelaskan pengaruh terjadinya keterlambatan mahasiswa dalam masalah studi akademik.

Teknik Analisis Data

Langkah- langkah analisis data yang dilakukan dalam penelitian ini terdiri dari beberapa tahapan sebagai berikut:

1. Mengumpulkan data mahasiswa teknik informatika universitas asahan dan data mahasiswa yang mengikuti organisasi seperti waktu dalam pembelajaran akademik dan keaktifan mahasiswa dalam mengikuti organisasi.
2. Menentukan dan menyiapkan dataset yang akan digunakan sebagai dasar analisis.
3. Melakukan prapemrosesan data, termasuk pembersihan data dan normalisasi.
4. Menerapkan algoritma gradient boosting untuk membentuk model pohon keputusan yang mengklasifikasikan mahasiswa dalam masalah akademik maupun non-akademik.
5. Mengevaluasi model menggunakan metrik evaluasi seperti akurasi dan presisi.
6. Menganalisis hasil klasifikasi untuk mengetahui hubungan antar akademik studi mahasiswa dan keaktifan mengikuti organisasi kemahasiswaan.

Teknik Analisis Sistem

Analisis sistem dalam penelitian ini mencakup perancangan dan implementasi sistem prediksi berbasis *machine learning*. Analisis sistem dilakukan menggunakan beberapa pendekatan sebagai berikut:

1. Sistem dapat melakukan klasifikasi data mahasiswa menggunakan algoritma machine learning berdasarkan data yang dimasukkan.
2. Sistem mendukung input data akademik mahasiswa dan perorganisasian secara manual maupun melalui file dataset yang dapat diproses langsung.
3. Sistem dapat menampilkan hasil klasifikasi dalam bentuk visualisasi pohon keputusan yang mudah dipahami.
4. Sistem diimplementasikan menggunakan bahasa pemrograman *Python 3.9* dengan menggunakan *library* utama seperti *pandas* dan *numpy* untuk manipulasi data, *scikit-learn* untuk pra-pemrosesan dan evaluasi model, *XGBoost* untuk implementasi algoritma *Gradient Boosting*, *imbalanced-learn* untuk penanganan ketidakseimbangan kelas, *matplotlib* dan *seaborn* untuk visualisasi data, serta *SHAP (SHapley Additive exPlanations)* untuk interpretasi model yang lebih mendalam.
5. Pengujian sistem dilakukan melalui Pengujian model menggunakan data testing yang telah disisihkan sebelumnya untuk mengukur performa generalisasi model dan Pengujian fungsional sistem untuk memastikan setiap komponen sistem bekerja dengan benar dan sesuai dengan kebutuhan yang telah didefinisikan. Hasil pengujian akan didokumentasikan dan dianalisis secara *komprehensif* dalam Bab 4.

3. Hasil dan Pembahasan

Deskripsi Dataset

Data yang digunakan dalam penelitian ini bersumber dari data alumni Program Studi S1 Teknik Informatika, Universitas Asahan, yang diperoleh dari Pangkalan Data Pendidikan Tinggi (PDDIKTI) per tanggal 17 Juni 2026. Dataset awal terdiri atas 28 baris data mahasiswa lulus, dengan atribut NIM, Nama, Jenis Kelamin, Status, Tanggal Keluar, Nomor Ijazah, Lama Pendidikan (dalam tahun), dan Indeks Prestasi Kumulatif (IPK).

Pada tahap praproses, ditemukan 3 baris data dengan nilai Lama Pendidikan dan/atau IPK kosong (ditandai "-"), yaitu mahasiswa dengan status kelulusan yang datanya belum lengkap tercatat pada sistem. Ketiga baris tersebut dikeluarkan dari proses pelatihan model, sehingga data bersih (clean data) yang digunakan untuk melatih model Gradient Boosting berjumlah 25 data.

Keterangan	Jumlah
Total data mentah (raw)	28
Data dikecualikan (lama studi / IPK kosong)	3
Data bersih digunakan untuk pelatihan	25

Pelabelan Data (Penentuan Variabel Target)

Penelitian ini memprediksi status keterlambatan studi mahasiswa, yaitu apakah seorang mahasiswa berpotensi "Tepat Waktu" atau "Terlambat" dalam menyelesaikan studi S1. Karena dataset historis hanya memuat data alumni (mahasiswa yang telah lulus), label kelas ditentukan secara terprogram berdasarkan lama studi aktual masing-masing alumni terhadap masa studi normal program S1, yaitu 4 tahun (8 semester):

label = "Terlambat" jika Lama Studi > 4 tahun; label = "Tepat Waktu" jika Lama Studi ≤ 4 tahun
 Dari 25 data bersih, diperoleh distribusi kelas sebagai berikut:

Kelas	Jumlah Data	Persentase
Tepat Waktu (≤ 4 tahun)	6	24,0%
Terlambat (> 4 tahun)	19	76,0%

Fitur prediktor (independent variable) yang digunakan adalah IPK dan Jenis Kelamin, karena kedua atribut tersebut tersedia secara konsisten pada data mahasiswa baik yang sudah lulus maupun yang masih aktif kuliah, sehingga model yang dilatih dapat diterapkan untuk memprediksi mahasiswa aktif sekalipun. Atribut Tahun Angkatan dan Lama Studi tidak digunakan sebagai fitur prediktor karena Lama Studi adalah sumber label itu sendiri (kebocoran data/data leakage), sedangkan Tahun Angkatan pada data historis alumni cenderung berkorelasi langsung dengan label (mahasiswa angkatan lama otomatis memiliki kesempatan lebih besar untuk "terlambat" dalam pencatatan historis), sehingga penggunaannya akan menyesatkan untuk prediksi mahasiswa angkatan baru yang masih aktif. Tabel 4.1 menyajikan keseluruhan data bersih yang digunakan sebagai data pelatihan (training set) model.

Tabel Data Alumni Hasil Praproses (Training Set)

No	NIM	Nama	JK	Angkatan	IPK	Lama (thn)	Label
1	18012001	Raja Aldi Zulkarnain	L	2018	3.35	6.4	Terlambat
2	18012005	Mhd. Bobby Iskandar	L	2018	3.33	6.4	Terlambat
3	18012034	Muhammad Nanda Azhari Tanjung	L	2018	3.12	6.4	Terlambat
4	18012039	Nanang Rinzani	L	2018	3.02	6.1	Terlambat
5	19012005	Ahmad Bustami	L	2019	3.53	5.9	Terlambat
6	19012014	Muhammad Reza Khatami Yusuf	L	2019	3.40	5.4	Terlambat
7	19012021	Amin Syahputra	L	2019	3.57	5.4	Terlambat
8	19012022	Marwan Belizato Zebua	L	2019	3.42	5.4	Terlambat
9	19012026	Reza Revo Fahlefi	L	2019	3.47	5.4	Terlambat
10	20012001	Fitri Handayani Lubis	P	2020	3.62	4.9	Terlambat
11	20012003	Rusli	L	2020	3.60	5.3	Terlambat
12	20012004	Kiki Rizki Rianda	P	2020	3.58	4.9	Terlambat
13	20012008	Dame Lasmaria Simangunsong	P	2020	3.22	4.9	Terlambat
14	20012010	Putri Indriani Sihombing	P	2020	3.65	4.9	Terlambat
15	20012011	Arisa Febriani	P	2020	3.44	4.9	Terlambat
16	20012022	Ilham Rinaldi	L	2020	3.38	5.3	Terlambat
17	20012025	Mairani Marpaung	P	2020	3.60	4.9	Terlambat
18	20012027	Desy Rahma Dani	P	2020	3.42	4.9	Terlambat
19	21012115	Ghifary Ridha Nasution	L	2021	2.91	4.0	Tepat Waktu
20	21012119	Nur Adzilla Aulia	P	2021	3.49	4.3	Terlambat
21	22012044	MHD KURNIAWAN	L	2022	3.26	3.3	Tepat Waktu

No	NIM	Nama	JK	Angkatan	IPK	Lama (thn)	Label
22	22012122	YUWANDA	P	2022	3.42	3.3	Tepat Waktu
23	22012154	DANDI IBNU ADITYA DAULAY	L	2022	3.28	3.3	Tepat Waktu
24	22012157	ASHYOKA INCA ARM	L	2022	3.11	3.3	Tepat Waktu
25	22012158	BETHARIA NOVALINA SITORUS	P	2022	3.64	3.3	Tepat Waktu

Implementasi Metode Gradient Boosting (XGBoost)

Bagian ini menjabarkan proses perhitungan metode Gradient Boosting (XGBoost) secara matematis menggunakan data riil hasil praproses pada subbab 4.1, mulai dari inisialisasi, pembentukan pohon (tree), hingga prediksi akhir. Pelatihan model dilakukan secara offline menggunakan pustaka XGBoost (Python) dengan parameter pada Tabel dibawa ini, dan hasilnya (struktur tree) diekspor untuk digunakan oleh aplikasi web (lihat subbab 4.3).

Tabel Parameter (Hyperparameter) Model

Parameter	Nilai	Keterangan
Fitur	ipk, gender	gender: 1 = Laki-laki (L), 0 = Perempuan (P)
n_estimators (jumlah tree)	5	jumlah iterasi boosting
learning_rate (η / eta)	0.3	laju kontribusi tiap tree terhadap prediksi akhir
max_depth	2	kedalaman maksimum tiap pohon
base_score (F_0)	0.5	prediksi awal sebelum boosting (probabilitas)
lambda (λ)	1.0	regularisasi L2 pada bobot leaf
gamma (γ)	0.0	ambang minimum gain agar split dilakukan
objective	binary:logistic	klasifikasi biner dengan fungsi logistik

Inisialisasi Prediksi Awal

Algoritma Gradient Boosting diawali dengan prediksi awal $F_0(x)$ yang sama untuk seluruh data, berasal dari $base_score = 0,5$. Karena fungsi objektif menggunakan skala log-odds (logit), $base_score$ ditransformasi ke margin awal:

$$F_0 = \ln(base_score / (1 - base_score)) = \ln(0,5 / 0,5) = 0$$

4.2.2 Perhitungan Gradien dan Hessian

Untuk fungsi loss logistik (binary log-loss), gradien (turunan pertama) dan hessian (turunan kedua) terhadap prediksi setiap sampel ke- i dirumuskan sebagai:

$$g_i = p_i - y_i \quad h_i = p_i (1 - p_i)$$

dengan $p_i = \text{sigmoid}(F_0) = 0,5$ untuk seluruh data pada iterasi pertama, dan y_i adalah label aktual (1 = Terlambat, 0 = Tepat Waktu). Sehingga setiap data memiliki $h_i = 0,5 \times 0,5 = 0,25$, dan g_i bernilai $-0,5$ untuk mahasiswa berlabel Terlambat, atau $+0,5$ untuk mahasiswa berlabel Tepat Waktu.

Pembentukan Pohon ke-1 (Tree #1)

Algoritma XGBoost mencari titik pemisah (split) terbaik pada fitur dengan menghitung nilai Gain dari setiap kandidat split. Untuk fitur IPK, kandidat split diuji pada titik tengah antar nilai IPK yang berurutan. Hasil pencarian menunjukkan split terbaik berada pada IPK = 3,33, dengan rumus Gain:

$$Gain = [G_l^2 / (H_l + \lambda)] + [G_r^2 / (H_r + \lambda)] - [(G_l + G_r)^2 / (H_l + H_r + \lambda)] - \gamma$$

Untuk split IPK < 3,33 (menyisakan 7 data di sisi kiri: IPK 2,91 s.d. 3,28, dengan 6 berlabel Tepat Waktu dan 1 Terlambat) dan IPK $\geq$ 3,33 (18 data di sisi kanan), diperoleh:

Sisi	n	Σ Gradien (G)	Σ Hessian (H)
Kiri (IPK < 3,33)	7	0,5	1,75
Kanan (IPK $\geq$ 3,33)	18	-7,0	4,5

Substitusi ke rumus Gain:

$$Gain = (0,5^2 / 2,75) + ((-7,0)^2 / 5,5) - ((-6,5)^2 / 7,25) = 0,0909 + 8,9091 - 5,8276 = 3,1724$$

Nilai Gain = 3,1724 ini merupakan kandidat split dengan gain tertinggi dibanding kandidat split lain (termasuk split pada fitur Jenis Kelamin), sehingga dipilih sebagai split akar (root) pada Tree #1. Karena $max_depth = 2$ dan

DOI: <https://doi.org/10.69693/ijmst.v4i2.11448>

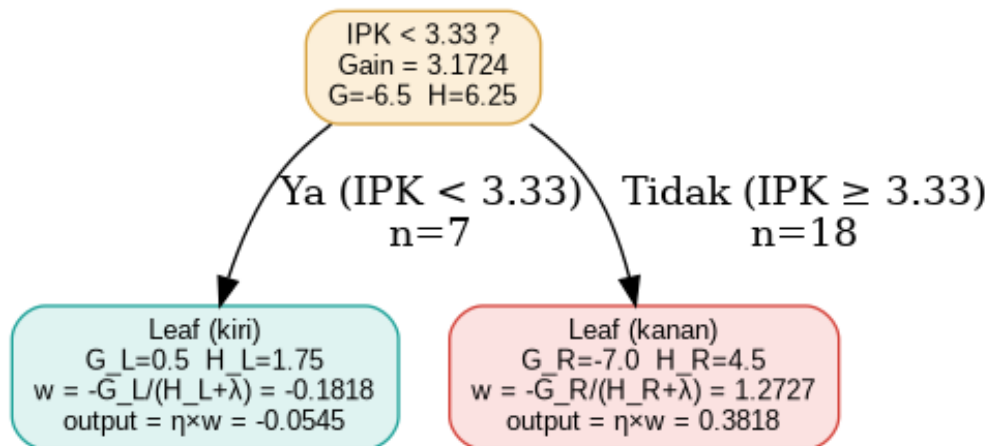
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

jumlah data pada cabang kiri sudah relatif homogen, algoritma tidak melakukan split lanjutan pada cabang ini (Gain split berikutnya di bawah ambang). Bobot (weight) tiap leaf dihitung dengan:

$$w^* = -G / (H + \lambda)$$

$$w \text{ kiri} = -0,5 / (1,75+1) = -0,1818 \rightarrow \text{output} = \eta \times w = 0,3 \times (-0,1818) = -0,0545$$

$$w \text{ kanan} = -7,0 / (4,5+1) = 1,2727 \rightarrow \text{output} = \eta \times w = 0,3 \times 1,2727 = 0,3818$$



Gambar Struktur Tree #1 hasil perhitungan manual (cocok dengan output pelatihan XGBoost)

Nilai Gain dan leaf weight hasil perhitungan manual di atas (Gain = 3,1724; leaf kiri = -0,0545; leaf kanan = 0,3818) telah diverifikasi identik dengan output asli pustaka XGBoost pada proses pelatihan, yang membuktikan kebenaran implementasi rumus yang digunakan.

Iterasi Pohon ke-2 sampai ke-5

Setelah Tree #1 terbentuk, prediksi (margin) setiap data diperbarui dengan menambahkan output Tree #1, kemudian gradien dan hessian dihitung ulang berdasarkan prediksi baru (residual yang tersisa). Proses ini berulang sebanyak $n_{\text{estimators}} = 5$ kali. Tabel merangkum hasil pembentukan tiap pohon:

Tree	Split Akar	Gain Akar	Leaf Kiri (IPK<3,33)	Cabang Kanan (IPK≥3,33)
#1	IPK < 3,33	3.1724	-0.0545	0.3818
#2	IPK < 3,33	1.9406	-0.0442	0.2979
#3	IPK < 3,33	1.2610	-0.0358	split Jenis Kelamin (Gain=0.1205); leaf P=0.1024, leaf L=0.3018
#4	IPK < 3,33	0.8967	-0.0290	split Jenis Kelamin (Gain=0.1176); leaf P=0.0842, leaf L=0.2643
#5	IPK < 3,33	0.6492	-0.0235	split Jenis Kelamin (Gain=0.1192); leaf P=0.0693, leaf L=0.2355

Pola yang konsisten terlihat pada seluruh pohon: split akar selalu terjadi pada fitur IPK dengan titik potong 3,33, sedangkan cabang kanan (IPK ≥ 3,33) pada Tree #3, #4, dan #5 mengalami split lanjutan pada fitur Jenis Kelamin. Hal ini menunjukkan IPK merupakan fitur paling berpengaruh (dominan) terhadap risiko keterlambatan studi pada dataset ini, dengan Jenis Kelamin berperan sebagai faktor penyeimbang sekunder.

Prediksi Akhir: Studi Kasus

Prediksi akhir suatu data x diperoleh dengan menjumlahkan margin awal F_0 dan kontribusi (output) dari seluruh tree, kemudian ditransformasi dengan fungsi sigmoid menjadi probabilitas:

$$\text{margin}(x) = F_0 + \sum_{m=1..5} \eta \cdot \text{leaf}_m(x) \quad P(\text{Terlambat} | x) = 1 / (1 + e^{-\text{margin}})$$

Studi Kasus 1 — Rusli (NIM 20012003, IPK 3,60, Laki-laki, label aktual: Terlambat)

Tree	Jejak Keputusan (path)	Leaf Output	Akumulasi Margin
#1	IPK ≥ 3,33 → kanan	0.381818	0.381818
#2	IPK ≥ 3,33 → kanan	0.297893	0.679711
#3	IPK ≥ 3,33 → kanan; JK ≥ 1 (L) → kanan	0.301798	0.981509
#4	IPK ≥ 3,33 → kanan; JK ≥ 1 (L) → kanan	0.264313	1.245823

Tree	Jejak Keputusan (path)	Leaf Output	Akumulasi Margin
#5	IPK $\geq 3,33 \rightarrow$ kanan; JK ≥ 1 (L) $\rightarrow$ kanan	0.235500	1.481323

$margin = 0 + 0,3818 + 0,2979 + 0,3018 + 0,2643 + 0,2355 = 1,4813$
 $P(\text{Terlambat}) = 1 / (1 + e^{-1,4813}) = 0,8148 \rightarrow 0,8148 \geq 0,5 \rightarrow \text{Hasil: TERLAMBAT (sesuai label aktual)}$
 Studi Kasus 2 — Ghifary Ridha Nasution (NIM 21012115, IPK 2,91, Laki-laki, label aktual: Tepat Waktu)

Tree	Jejak Keputusan (path)	Leaf Output	Akumulasi Margin
#1	IPK $< 3,33 \rightarrow$ kiri	-0.054545	-0.054545
#2	IPK $< 3,33 \rightarrow$ kiri	-0.044156	-0.098701
#3	IPK $< 3,33 \rightarrow$ kiri	-0.035773	-0.134474
#4	IPK $< 3,33 \rightarrow$ kiri	-0.028995	-0.163469
#5	IPK $< 3,33 \rightarrow$ kiri	-0.023507	-0.186976

$margin = 0 - 0,0545 - 0,0442 - 0,0358 - 0,0290 - 0,0235 = -0,1870$
 $P(\text{Terlambat}) = 1 / (1 + e^{-(-0,1870)}) = 0,4534 \rightarrow 0,4534 < 0,5 \rightarrow \text{Hasil: TEPAT WAKTU (sesuai label aktual)}$

Kedua studi kasus di atas menunjukkan bahwa hasil perhitungan manual sepenuhnya identik dengan hasil yang ditampilkan oleh aplikasi (lihat Gambar 4.7 pada subbab 4.4), yang membuktikan implementasi ulang algoritma XGBoost pada sisi PHP (tanpa pustaka eksternal) telah dilakukan secara benar dan konsisten dengan model aslinya.

Evaluasi Model

Evaluasi performa model dilakukan dengan menghitung confusion matrix pada seluruh 25 data pelatihan, dengan ambang keputusan probabilitas 0,5:

	Prediksi: Tepat Waktu	Prediksi: Terlambat
Aktual: Tepat Waktu	4 (TN)	2 (FP)
Aktual: Terlambat	3 (FN)	16 (TP)

$Akurasi = (TN+TP)/n = 80.0\%$
 $Presisi = TP/(TP+FP) = 88.9\%$
 $Recall = TP/(TP+FN) = 84.2\%$
 $F1\text{-Score} = 2 \times (Presisi \times Recall) / (Presisi + Recall) = 86.5\%$

Model memperoleh akurasi 80,0% pada data pelatihan, dengan 2 kesalahan klasifikasi jenis False Positive (mahasiswa aktual Tepat Waktu diprediksi Terlambat) dan 3 kesalahan False Negative (mahasiswa aktual Terlambat diprediksi Tepat Waktu) dari total 25 data. Mengingat jumlah data yang relatif kecil (25 data) dan ketidakseimbangan kelas (76% berlabel Terlambat), hasil ini cukup baik sebagai purwarupa (prototype), namun pembahasan lebih lanjut mengenai keterbatasan ini disajikan pada subbab 4.6.

Perancangan dan Implementasi Sistem

Aplikasi "SIPKS - Sistem Prediksi Keterlambatan Studi" dibangun sebagai aplikasi web menggunakan PHP (native, tanpa framework) dengan database MySQL/MariaDB dan antarmuka berbasis Bootstrap 5. Subbab ini menjelaskan tahapan pembangunan aplikasi, arsitektur, alur kerja sistem, serta rancangan basis data.

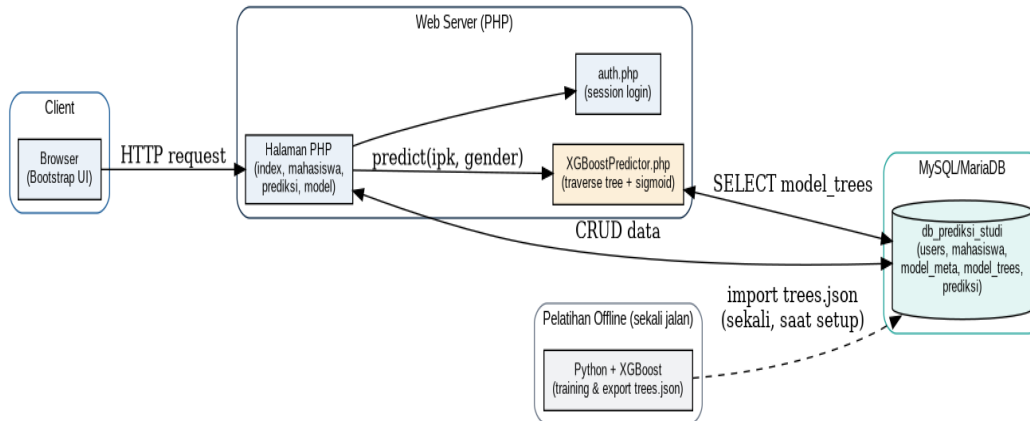
Tahapan Pembangunan Aplikasi

1. Analisis kebutuhan: menentukan variabel target (status keterlambatan studi), fitur prediktor (IPK, Jenis Kelamin), dan pengguna sistem (admin/operator akademik).
2. Pengumpulan dan praproses data: ekspor data alumni dari PDDIKTI, pembersihan data kosong/tidak valid, dan pelabelan kelas berdasarkan ambang masa studi (lihat subbab 4.1).
3. Pelatihan model secara offline: pelatihan model Gradient Boosting (XGBoost) menggunakan Python dan pustaka xgboost pada data hasil praproses, kemudian mengekspor struktur pohon (trees) dan parameter model ke format JSON.
4. Perancangan basis data: merancang skema tabel untuk menyimpan data mahasiswa, struktur model (trees), serta riwayat prediksi (lihat subbab 4.3.3).
5. Implementasi mesin inferensi (XGBoostPredictor.php): menerjemahkan ulang logika penelusuran pohon dan fungsi sigmoid XGBoost ke dalam PHP murni, agar prediksi dapat dijalankan langsung di server web tanpa dependensi pustaka machine learning eksternal.

6. Implementasi antarmuka (UI/UX): membangun halaman login, dashboard, manajemen data mahasiswa, form prediksi, halaman rincian perhitungan, riwayat, dan informasi model menggunakan Bootstrap 5 agar tampilan responsif dan mudah digunakan (user friendly).
7. Pengujian sistem: pengujian fungsional (black box) terhadap seluruh alur aplikasi serta verifikasi kebenaran hasil perhitungan dibandingkan dengan output model Python asli (lihat subbab 4.2.5 dan 4.5).
8. Evaluasi dan penyempurnaan: perbaikan antarmuka, penanganan kesalahan input, serta penambahan fitur riwayat dan transparansi perhitungan (explainability).

Arsitektur Sistem

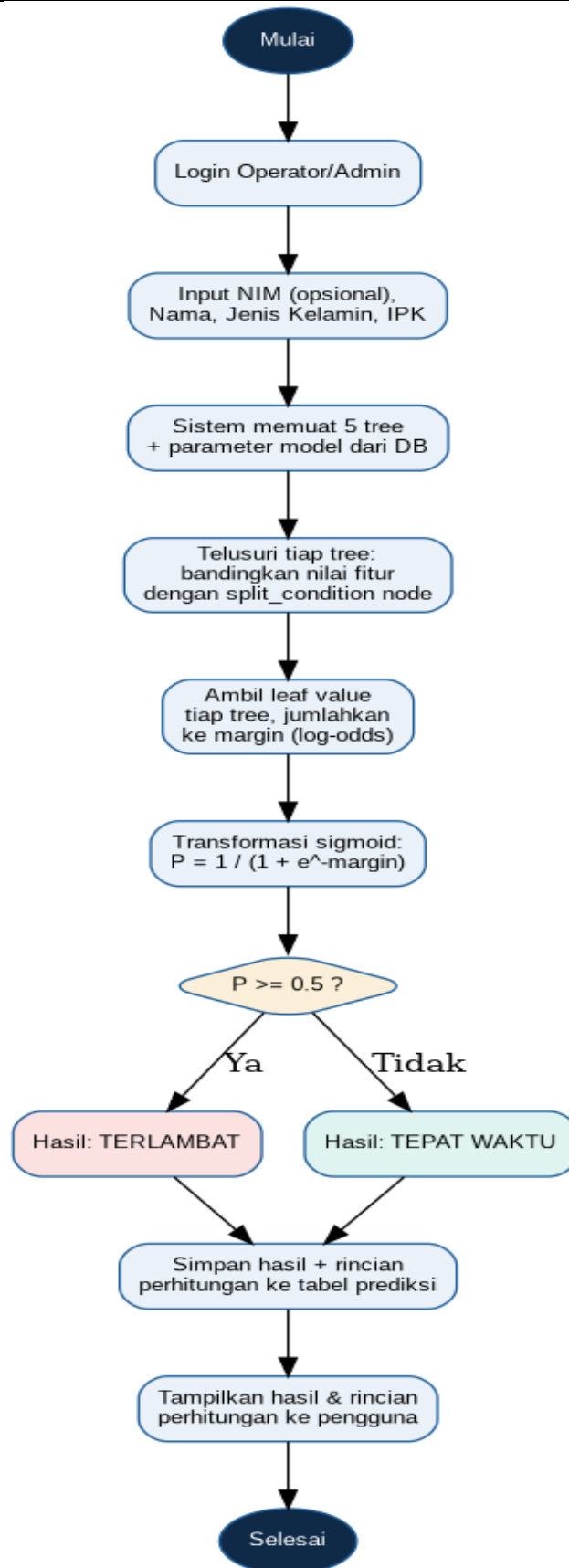
Sistem terdiri atas tiga lapisan utama: (1) antarmuka pengguna berbasis browser dengan Bootstrap, (2) lapisan aplikasi PHP yang menangani autentikasi, manajemen data, dan mesin inferensi XGBoostPredictor, serta (3) basis data MySQL/MariaDB yang menyimpan seluruh data dan struktur model. Proses pelatihan model dilakukan terpisah (offline) menggunakan Python dan hasilnya diimpor satu kali ke basis data saat instalasi/setup sistem, sehingga server produksi tidak memerlukan instalasi pustaka machine learning.



Gambar Arsitektur Sistem SIPKS

Alur Kerja Sistem (Workflow)

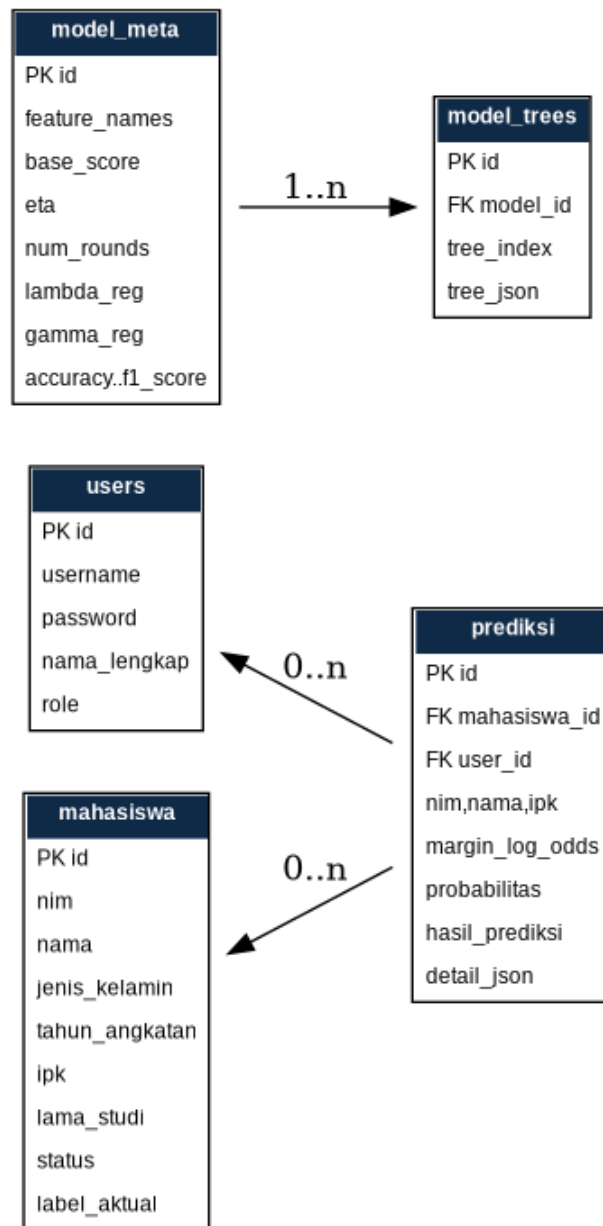
Gambar 4.3 menunjukkan alur kerja sistem secara keseluruhan, mulai dari proses login, input data, hingga proses inferensi model dan penyimpanan riwayat hasil prediksi.



Gambar 4.3 Diagram Alur Kerja (Flowchart) Sistem

Perancangan Basis Data

Basis data dirancang dengan lima tabel utama: users (akun login), mahasiswa (data historis alumni/mahasiswa aktif), model_meta (parameter dan metrik evaluasi model), model_trees (struktur pohon hasil pelatihan, dalam format JSON), dan prediksi (riwayat hasil prediksi beserta rincian perhitungannya). Relasi antartabel ditunjukkan pada Gambar



Gambar Entity Relationship Diagram (ERD) Basis Data

Struktur tabel utama dijabarkan pada Tabel dan Tabel.

Tabel Struktur Tabel mahasiswa

Kolom	Tipe Data	Keterangan
id	INT, PK, AUTO_INCREMENT	Kunci utama
nim	VARCHAR(20), UNIQUE	Nomor Induk Mahasiswa
nama	VARCHAR(150)	Nama lengkap mahasiswa
jenis_kelamin	ENUM('L','P')	Jenis kelamin
tahun_angkatan	SMALLINT	Tahun masuk

Kolom	Tipe Data	Keterangan
ipk	DECIMAL(3,2)	Indeks Prestasi Kumulatif
lama_studi	DECIMAL(4,2), nullable	Lama studi (tahun), kosong jika masih aktif
status	ENUM('Aktif','Lulus','DO')	Status mahasiswa
label_aktual	ENUM('Tepat Waktu','Terlambat')	Label aktual (hanya terisi utk alumni)

Tabel Struktur Tabel prediksi

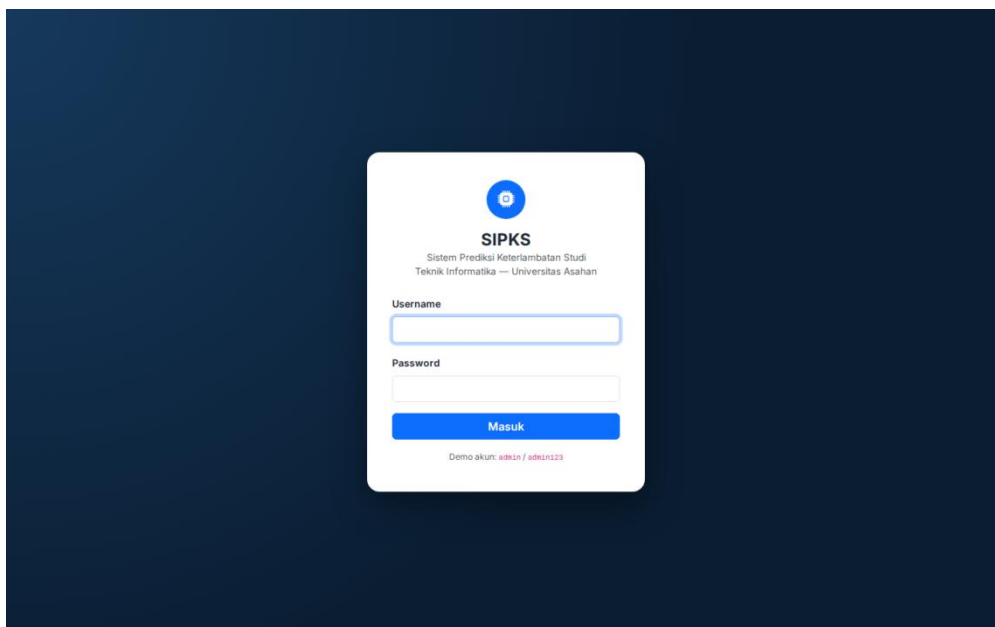
Kolom	Tipe Data	Keterangan
id	INT, PK, AUTO_INCREMENT	Kunci utama
mahasiswa_id	INT, FK, nullable	Relasi ke tabel mahasiswa (jika NIM dikenali)
nim, nama, jenis_kelamin, ipk	VARCHAR/DECIMAL	Data input prediksi
margin_log_odds	DECIMAL(10,6)	Hasil margin sebelum sigmoid
probabilitas	DECIMAL(6,4)	Probabilitas kelas Terlambat (0-1)
hasil_prediksi	ENUM('Tepat Waktu','Terlambat')	Kelas hasil prediksi
detail_json	LONGTEXT	Rincian jejak perhitungan tiap tree (utk ditampilkan di UI)
user_id	INT, FK, nullable	Operator yang menjalankan prediksi
created_at	TIMESTAMP	Waktu prediksi dijalankan

Implementasi Antarmuka

Antarmuka aplikasi dibangun responsif menggunakan Bootstrap 5, dengan tata letak sidebar-topbar yang konsisten di seluruh halaman, serta skema warna navy-teal-gold yang mencerminkan identitas akademik. Berikut ditampilkan hasil implementasi tiap halaman utama.

Halaman Login

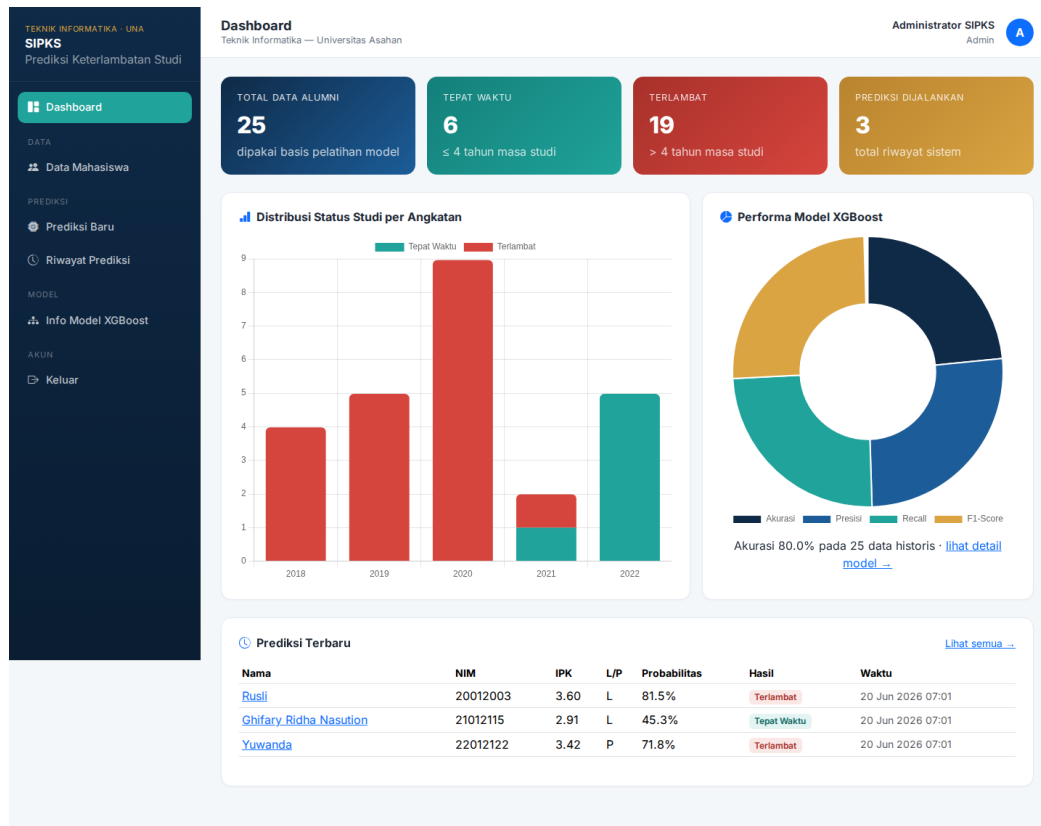
Halaman login digunakan untuk autentikasi admin/operator sebelum dapat mengakses sistem. Kata sandi disimpan dalam bentuk hash bcrypt menggunakan fungsi password_hash() PHP.



Gambar Halaman Login

Dashboard

Dashboard menampilkan ringkasan statistik data alumni (total data, jumlah Tepat Waktu/Terlambat), grafik distribusi status studi per angkatan, grafik performa model (akurasi, presisi, recall, F1-score), serta daftar prediksi terbaru.



Gambar Halaman Dashboard

Manajemen Data Mahasiswa

Halaman ini menyediakan fungsi CRUD (Create, Read, Update, Delete) terhadap data mahasiswa/alumni yang menjadi basis pelatihan model, lengkap dengan pencarian berdasarkan NIM>Nama serta label otomatis Tepat Waktu/Terlambat.

TEKNIK INFORMATIKA - UNA

SIPKS

Prediksi Keterlambatan Studi

Dashboard

DATA

Data Mahasiswa

PREDIKSI

Prediksi Baru

Riwayat Prediksi

MODEL

Info Model XGBoost

AKUN

Keluar

Data Mahasiswa / Alumni

Teknik Informatika — Universitas Asahan

Administrator SIPKS

Admin

Tambah Data Mahasiswa

NIM

Nama Lengkap

Jenis Kelamin

Laki-laki

Tahun Angkatan

2026

IPK

Lama Studi (thn)

kosongkan jika aki

Status

Aktif

Label "Tepat Waktu / Terlambat" otomatis dihitung dari Lama Studi (ambang < 4 tahun).

Tambah Data

Daftar Mahasiswa (25)

Cari NIM/Nama...

NIM	Nama	L/P	Angkatan	IPK	Lama Studi	Status	Label			
22012157	ASHYOKA INCA ARM	L	2022	3.11	3.3 thn	Lulus	Tepat Waktu			
22012158	BETHARIA NOVALINA SITORUS	P	2022	3.64	3.3 thn	Lulus	Tepat Waktu			
22012154	DANDI IBNU ADITYA DAULAY	L	2022	3.28	3.3 thn	Lulus	Tepat Waktu			
22012044	MHD KURNIAWAN	L	2022	3.26	3.3 thn	Lulus	Tepat Waktu			
22012122	YUWANDA	P	2022	3.42	3.3 thn	Lulus	Tepat Waktu			
21012115	Ghifary Ridha Nasution	L	2021	2.91	4.0 thn	Lulus	Tepat Waktu			
21012119	Nur Adzilla Aulia	P	2021	3.49	4.3 thn	Lulus	Terlambat			
20012011	Arisa Febriani	P	2020	3.44	4.9 thn	Lulus	Terlambat			
20012008	Dame Lasmaria Simangunsong	P	2020	3.22	4.9 thn	Lulus	Terlambat			
20012027	Desy Rahma Dani	P	2020	3.42	4.9 thn	Lulus	Terlambat			
20012001	Fitri Handayani Lubis	P	2020	3.62	4.9 thn	Lulus	Terlambat			
20012022	Ilham Rinaldi	L	2020	3.38	5.3 thn	Lulus	Terlambat			
20012004	Kiki Rizki Rianda	P	2020	3.58	4.9 thn	Lulus	Terlambat			
20012025	Mairani Marpaung	P	2020	3.60	4.9 thn	Lulus	Terlambat			
20012010	Putri Indriani Sihombing	P	2020	3.65	4.9 thn	Lulus	Terlambat			
20012003	Rusli	L	2020	3.60	5.3 thn	Lulus	Terlambat			
19012005	Ahmad Bustami	L	2019	3.53	5.9 thn	Lulus	Terlambat			
19012021	Amin Syahputra	L	2019	3.57	5.4 thn	Lulus	Terlambat			
19012022	Marwan Belizato Zebua	L	2019	3.42	5.4 thn	Lulus	Terlambat			
19012014	Muhammad Reza Khatami Yusuf	L	2019	3.40	5.4 thn	Lulus	Terlambat			
19012026	Reza Revo Fahlefi	L	2019	3.47	5.4 thn	Lulus	Terlambat			
18012005	Mhd. Bobby Iskandar	L	2018	3.33	6.4 thn	Lulus	Terlambat			
18012034	Muhammad Nanda Azhari Tanjung	L	2018	3.12	6.4 thn	Lulus	Terlambat			
18012039	Nanang Rinzani	L	2018	3.02	6.1 thn	Lulus	Terlambat			
18012001	Raja Aldi Zulkarnain	L	2018	3.35	6.4 thn	Lulus	Terlambat			

Gambar Halaman Data Mahasiswa/Alumni

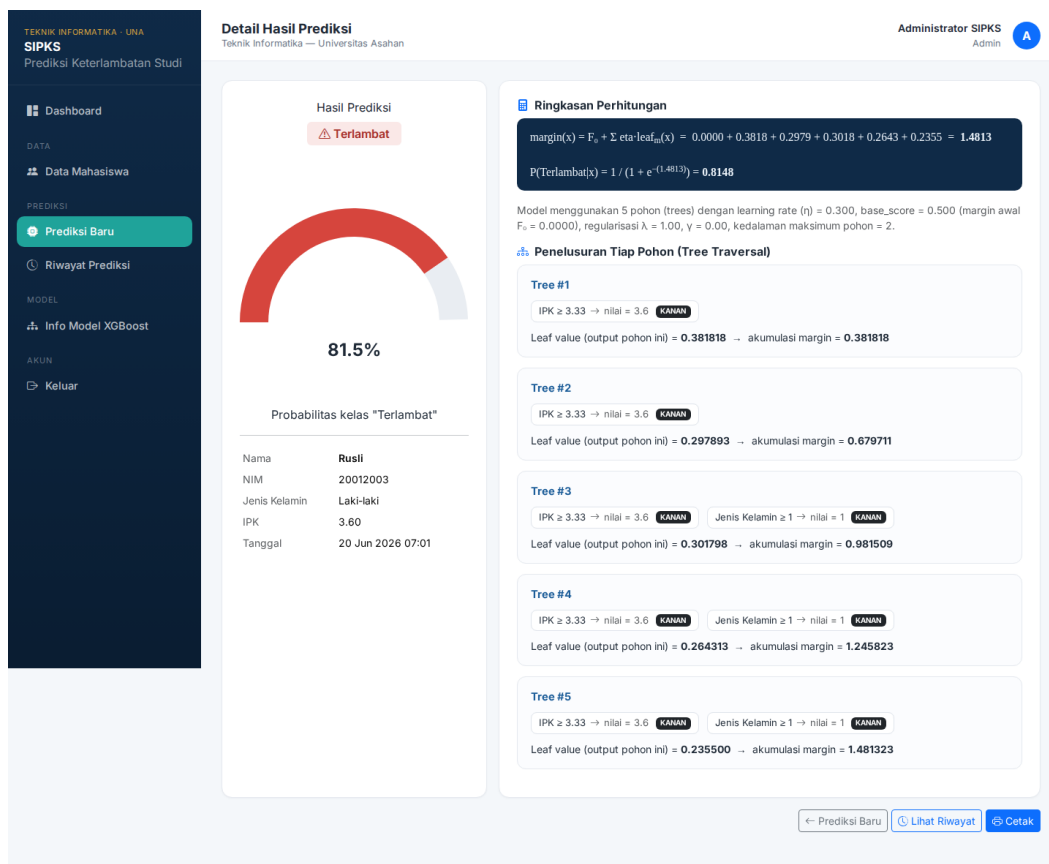
Form Prediksi

Pengguna memasukkan NIM (opsional), nama, jenis kelamin, dan IPK mahasiswa yang ingin diprediksi. Halaman ini juga menjelaskan secara ringkas mekanisme kerja model kepada pengguna.

Gambar Halaman Form Prediksi Baru

Detail Hasil Perhitungan

Halaman ini merupakan implementasi langsung dari proses perhitungan pada subbab 4.2: menampilkan hasil akhir (badge Tepat Waktu/Terlambat), gauge probabilitas, ringkasan rumus margin dan sigmoid, serta rincian penelusuran tiap pohon (tree traversal) secara transparan. Data pada Gambar 4.9 merupakan hasil prediksi untuk Studi Kasus 1 (Rusli) pada subbab 4.2.5, dengan hasil probabilitas 81,5% (0,8148 dibulatkan) yang identik dengan perhitungan manual.



Gambar Halaman Detail Hasil Prediksi (Studi Kasus Rusli)

Riwayat Prediksi

Seluruh prediksi yang pernah dijalankan tercatat dan dapat difilter berdasarkan hasil (Tepat Waktu/Terlambat), sehingga riwayat keputusan sistem dapat ditelusuri kembali sewaktu-waktu.

#	Nama	NIM	L/P	IPK	Margin	Probabilitas	Hasil	Operator	Waktu
1	Rusli	20012003	L	3.60	1.4813	81.5%	Terlambat	Administrator SIPKS	20 Jun 2026 07:01
2	Ghifary Ridha Nasution	21012115	L	2.91	-0.1870	45.3%	Tepat Waktu	Administrator SIPKS	20 Jun 2026 07:01
3	Yuwanda	22012122	P	3.42	0.9356	71.8%	Terlambat	Administrator SIPKS	20 Jun 2026 07:01

Gambar 4.10 Halaman Riwayat Prediksi

Informasi Model

Halaman ini menampilkan seluruh hyperparameter model, hasil evaluasi (confusion matrix dan metrik), serta struktur mentah (raw JSON) tiap pohon yang dapat diperluas (expand) untuk keperluan audit/transparansi model.

TEKNIK INFORMATIKA - UNA

SIPKS

Prediksi Keterlambatan Studi

Dashboard

DATA

Data Mahasiswa

PREDIKSI

Prediksi Baru

Riwayat Prediksi

MODEL

Info Model XGBoost

AKUN

Keluar

Info Model XGBoost

Teknik Informatika — Universitas Asahan

Administrator SIPKS

Admin

Hyperparameter Model

Fitur (features)	ipk,gender
Jumlah pohon (n_estimators)	5
Learning rate (η)	0.300
Kedalaman maksimum (max_depth)	2
Base score (F ₀)	0.500
Regularisasi L2 (λ)	1.00
Regularisasi kompleksitas (γ)	0.00
Dilatih pada	20 Jun 2026 07:01

Evaluasi Model (Confusion Matrix)

Akurasi	Presisi	Recall	F1-Score
80.0%	88.9%	84.2%	86.5%
Prediksi: Tepat Waktu			
Aktual: Tepat Waktu	4 (TN)	2 (FP)	
Prediksi: Terlambat			
Aktual: Terlambat	3 (FN)	16 (TP)	

Evaluasi dihitung pada seluruh 25 data alumni historis (training set) yang tersedia.

Struktur Pohon (Trees) Tersimpan

Sistem menyimpan struktur internal tiap pohon hasil pelatihan XGBoost (split feature, split condition, gain, leaf weight) dalam format JSON pada tabel `model_trees`, dan menelusurinya secara langsung di PHP saat prediksi dijalankan.

Tree #1	▼
Tree #2	▼
Tree #3	▼
Tree #4	▼
Tree #5	▼

Gambar Halaman Informasi Model XGBoost

Pengujian Sistem

Pengujian dilakukan dengan metode black box testing untuk memastikan setiap fungsi sistem berjalan sesuai rancangan, serta pengujian kesesuaian hasil perhitungan untuk memastikan akurasi implementasi algoritma.

Tabel Hasil Pengujian Black Box

No	Skenario Pengujian	Hasil yang Diharapkan	Status
1	Login dengan username/password benar	Berhasil masuk ke dashboard	Sesuai

No	Skenario Pengujian	Hasil yang Diharapkan	Status
2	Login dengan password salah	Menampilkan pesan kesalahan, tetap di halaman login	Sesuai
3	Tambah data mahasiswa baru	Data tersimpan dan tampil di daftar	Sesuai
4	Tambah data dengan NIM duplikat	Menampilkan pesan "NIM sudah terdaftar"	Sesuai
5	Edit & hapus data mahasiswa	Data ter-update/terhapus dari daftar dan database	Sesuai
6	Jalankan prediksi dengan IPK & JK valid	Sistem menampilkan hasil, probabilitas, dan rincian tree	Sesuai
7	Jalankan prediksi dengan IPK kosong/di luar rentang	Menampilkan pesan validasi, prediksi tidak diproses	Sesuai
8	Verifikasi hasil prediksi vs perhitungan manual/Python	Probabilitas identik (selisih 0)	Sesuai
9	Lihat riwayat & filter hasil prediksi	Daftar riwayat tampil sesuai filter yang dipilih	Sesuai
10	Akses halaman tanpa login (belum autentikasi)	Diarahkan otomatis ke halaman login	Sesuai
11	Logout	Sesi berakhir, diarahkan ke halaman login	Sesuai

Pengujian kesesuaian numerik (No. 8) dilakukan dengan menjalankan dua skenario data (Studi Kasus Rusli dan Ghifary, subbab 4.2.5) baik secara manual, melalui skrip Python (model asli), maupun melalui antarmuka aplikasi PHP. Ketiganya menghasilkan nilai margin dan probabilitas yang identik hingga empat angka di belakang koma, yang membuktikan implementasi ulang (re-implementation) algoritma XGBoost pada PHP bebas dari kesalahan perhitungan (calculation error).

Pembahasan

Kelebihan Sistem

1. Transparansi perhitungan: setiap hasil prediksi disertai rincian penelusuran tiap pohon, sehingga pengguna (misalnya dosen Pembimbing Akademik) dapat memahami dasar keputusan model, bukan sekadar label akhir (explainable AI).
2. Ringan dan portabel: mesin inferensi diimplementasikan murni dalam PHP tanpa dependensi pustaka machine learning eksternal di server produksi, sehingga mudah di-deploy pada hosting/shared server kampus yang umumnya hanya mendukung PHP dan MySQL.
3. Antarmuka responsif dan mudah digunakan, dengan dashboard ringkasan, riwayat, serta manajemen data terintegrasi dalam satu sistem.
4. Akurasi hasil terverifikasi identik dengan model XGBoost asli, sehingga kepercayaan terhadap hasil inferensi PHP terjamin.

Keterbatasan

1. Jumlah data pelatihan masih kecil (25 data) dan tidak seimbang (76% berlabel Terlambat), sehingga model berisiko bias terhadap kelas mayoritas dan generalisasinya terbatas pada data baru.
2. Fitur prediktor hanya mencakup IPK dan Jenis Kelamin karena keterbatasan atribut pada data historis yang tersedia. Penambahan fitur seperti IPS per semester, jumlah SKS lulus, status cuti akademik, dan jalur masuk berpotensi meningkatkan akurasi model secara signifikan.
3. Model belum dilatih ulang secara otomatis (retraining) di dalam aplikasi; pelatihan ulang masih harus dilakukan secara manual melalui skrip Python apabila data bertambah signifikan.
4. Pengujian evaluasi dilakukan pada data pelatihan itu sendiri (tanpa data uji terpisah/hold-out) karena keterbatasan jumlah data, sehingga metrik akurasi yang dilaporkan berpotensi optimistik (overfitting) dan perlu divalidasi lebih lanjut dengan data yang lebih besar.

Rekomendasi Pengembangan Lanjutan

1. Mengumpulkan data mahasiswa aktif secara berkala (per semester) untuk memperbesar dan menyeimbangkan dataset pelatihan.
2. Menambahkan fitur prediktor yang lebih kaya, misalnya IPS per semester, jumlah SKS tertunda, dan status keaktifan organisasi/kerja paruh waktu.
3. Menambahkan modul pelatihan ulang (retraining) otomatis berbasis penjadwalan (cron job) yang memanggil skrip Python dan memperbarui tabel model trees secara berkala.

4. Menerapkan validasi silang (k-fold cross validation) dan pemisahan data latih/uji (train-test split) seiring bertambahnya volume data, untuk mendapatkan estimasi performa model yang lebih objektif.

4. Kesimpulan

1. Model prediksi keterlambatan studi mahasiswa menggunakan algoritma *XGBoost* menunjukkan bahwa variabel IPK merupakan faktor paling dominan dalam menentukan status kelulusan, dengan titik pemisah optimal pada nilai IPK 3,33.
2. Hasil evaluasi model menunjukkan performa yang cukup baik dengan akurasi sebesar 80%, presisi 88,9%, recall 84,2%, dan F1-score 86,5%, sehingga model dinilai mampu melakukan klasifikasi keterlambatan studi secara efektif meskipun menggunakan jumlah data yang terbatas.
3. Implementasi sistem berbasis web (SIPKS) menggunakan PHP dan MySQL berhasil menerapkan model *XGBoost* tanpa pustaka eksternal, serta menghasilkan prediksi yang konsisten dengan hasil perhitungan manual dan pelatihan model, sehingga sistem dapat digunakan sebagai alat bantu pengambilan keputusan akademik.
4. Meskipun model menunjukkan hasil yang baik, keterbatasan jumlah data hanya 25 data dan ketidakseimbangan kelas masih mempengaruhi tingkat generalisasi model, sehingga hasil masih bersifat sebagai prototipe awal.

Reference

- Agusmawati, N. K., Arifin, Z., & Tholib, A. (2024). Implementasi Algoritma Machine Learning Bidang Pendidikan Berbasis Web Menggunakan Framework Streamlit. Brandedtech Publisher. <https://books.google.co.id/books?id=xEvGEQAAQBAJ>
- Al-Husaini, M., Saputra, P. A., Renaldi, M., Maulana, R. A., Efitra, E., & Permata, N. G. (2024). Prediksi Tsunami Dengan Metode Ensemble Machine Learning. PT. Sonpedia Publishing Indonesia. <https://books.google.co.id/books?id=QAJtEAAAQBAJ>
- Alam, P., Tanjungpura, U., & Mining, D. (2024). Gradient Boosting Machine pada Klasifikasi Kelulusan Mahasiswa Virginia Atlantic, Evy Sulistianingsih, Hendra Perdana. Jurnal Data Mining, 13(2), 165–174.
- Arbin Janu Setiyowati. (2023). Academic Burnout Siswa dan Implikasinya Terhadap Layanan Bimbingan dan Konseling di Sekolah. Media Nusa Creative (MNC Publishing). <https://books.google.co.id/books?id=sY3PEAAAQBAJ>
- Dahlan Abdullah. (2026). Kecerdasan Komputasional. Penerbit KBM Indonesia. <https://books.google.co.id/books?id=WBzPEQAAQBAJ>
- Fadhillah, M. B. A., Anggoro, R., & Priyatno, A. M. (2020). Analisis Performa Link Stability dari Faktor Kecepatan untuk Dinamisasi Zona pada Zone Routing Protocol. Jurnal Rekayasa Elektrika, 16(3). <https://doi.org/10.17529/jre.v16i3.16502>
- Harefa, E., & Afendi, H. (2024). Buku Ajar Teori Belajar dan Pembelajaran. PT. Sonpedia Publishing Indonesia. <https://books.google.co.id/books?id=XCT2EAAAQBAJ>
- Hartono. (n.d.). Modul digital machine learning.
- Herjanto, E. (2019). Manajemen Operasi (Edisi 3). Grasindo. <https://books.google.co.id/books?id=xGgDqd15N2EC>
- Imron, S., Faizah, A., & Informasi, F. T. (2026). Prediksi Kelulusan Mahasiswa Menggunakan Algoritma XGBoost. Jurnal Teknologi Informasi, 9, 76–86.
- Iyan. (2024). Pengantar Ilmu Komputer (Issue 67).
- Jamhuri, M., Irawan, M. I., Imam Mukhlash, A. K., & Adab, P. (n.d.). Buku Panduan Praktis Machine Learning dengan Python. Penerbit Adab. <https://books.google.co.id/books?id=dNLGEQAAQBAJ>
- Kadang, M. O., Indahsari, A. N., Sumah, J., Firdaus, F., Nasution, V. M., & Prayoga, J. (2025). Machine Learning Tanpa Label: Panduan Lengkap Unsupervised Learning. Serasi Media Teknologi. <https://books.google.co.id/books?id=r3-XEQAAQBAJ>
- Mesran, S. (2020). Pengantar Teknologi Informasi. Graha Mitra Edukasi. <https://books.google.co.id/books?id=gK66EAAAQBAJ>
- Mira, S. K., & Azriel Christian Nurcahyo, S. K. M. (2025). Data Mining Mengeksplorasi Teknik-Teknik Data Mining dan Metode K-Means Teori, Konsep, Algoritma dan Studi Kasus. Uwais Inspirasi Indonesia. <https://books.google.co.id/books?id=32BQEQAQBAJ>
- Moh. Saiful Hakiki, drh. Dwi Aprilia Anggraini, Riyan Sisiawan Putra, Denis Fidita Karya, & Mohamad Rijal Iskandar Zhulqurnain. (2023). Buku Ajar Manajemen Keuangan 1A Konsep-Konsep Dan Studi Kasus Dalam Bidang Keuangan. Uwais Inspirasi Indonesia. <https://books.google.co.id/books?id=K6DeEAAAQBAJ>
- Mulyo, H., & Zyen, A. K. (2025). Bulletin of Computer Science Research Pengaruh Hyperparameter Tuning Gradient Boosting Terhadap Prediksi Pemilihan Program Studi Mahasiswa Baru. Bulletin of Computer Science Research, 5(2), 131–137. <https://doi.org/10.47065/bulletincsr.v5i2.454>
- Natekin, & Knoll. (2013). Gradient Boosting Machines. Deepublish. <https://books.google.co.id/books?id=eZRYEQAAQBAJ>
- Nst, E. N., Nurcahyo, G. W., Informatika, M. T., Begalung, L., & Barat, S. (2025). Penerapan Machine Learning Menggunakan Algoritma Decision Tree Untuk Prediksi. Jurnal Informatika, 8(2), 428–439.
- Prasetya, M. R. A., & Priyatno, A. M. (2022). Dice Similarity and TF-IDF for New Student Admissions Chatbot. RIGGS: Journal of Artificial Intelligence and Digital Business, 1(1), 13–18. <https://doi.org/10.31004/riggs.v1i1.5>
- Prof. Darmawan Utomo. (2025). Deep Learning Dasar Teori dan Implementasi dengan Python dan Keras. Diandra Kreatif (Penalicious). <https://books.google.co.id/books?id=nqmQEQAQBAJ>
- Priyatno, A. M. (2019). The Application of HAAR Wavelet and Backpropagation for Diabetic Retinopathy Classification Based on Eye Retina Image. International Journal of Science, Engineering, and Information Technology, 03(02), 139–142. <https://journal.trunojoyo.ac.id/ijseit/article/view/4536>
- Priyatno, A. M. (2020). Spammer Detection Based on Account, Tweet, and Community Activity on Twitter. Jurnal Ilmu Komputer dan Informasi, 13(2), 97–107. <https://doi.org/10.21609/jiki.v13i2.871>
- Priyatno, A. M., & Firmananda, F. I. (2022). N-Gram Feature for Comparison of Machine Learning Methods on Sentiment in Financial News Headlines. RIGGS: Journal of Artificial Intelligence and Digital Business, 1(1), 01–06. <https://doi.org/10.31004/riggs.v1i1.4>
- Priyatno, A. M., Muttaqi, M. M., Syuhada, F., & Arifin, A. Z. (2019). Deteksi bot spammer twitter berbasis time interval entropy dan global vectors for word representations tweet's hashtag. Register: Jurnal Ilmiah Teknologi Sistem Informasi, 5(1), 37–46. <https://doi.org/10.26594/register.v5i1.1382>

-
- Rachman, M. A., Maulana, D. R., Timothy, B., Annisa, R., Informatika, S., Bina, U., Informatika, S., Mining, D., & Forest, R. (2026). Aplikasi Prediksi Tingkat Kelulusan Mahasiswa Berdasarkan Data Akademik dan Demografi Menggunakan Algoritma Klasifikasi Random. *Jurnal Media Informatika (JUMIN)*, 7(1), 247–255.
- Ramadhan, W., & Trisnawan, A. B. (2026). Machine Learning Terapan: Prediksi Bisnis Dan Marketing Intelligence. <https://books.google.co.id/books?id=UznbEQAAQBAJ>
- Sanjaya, S., Priyatno, A. M., Yanto, F., & Afrianty, I. (2018). Klasifikasi Diabetik Retinopati Menggunakan Wavelet Haar dan Backpropagation Neural Network. *Seminar Nasional Teknologi Informasi Komunikasi Dan Industri (SNTIKI-10)*, 77–84.
- Sarwandi, L. (2023). Sistem Pendukung Keputusan. *Graha Mitra Edukasi*. <https://books.google.co.id/books?id=qmm-EAAAQBAJ>
- Septiani, S., Nababan, H. S., & Siregar, R. W. (2024). Manajemen Pendidikan Tinggi. Sada Kurnia Pustaka. <https://books.google.co.id/books?id=KDL3EAAAQBAJ>
- Sitorus, S. A., & Hasibuan, A., dkk. (2022). Pengantar Perilaku Organisasi.
- Suryantara, I. G. N. (2024). PYTHON: Implementasi Algoritma Kompleks dalam Era Industri 5.0 dan Society 5.0. Elex Media Komputindo. <https://books.google.co.id/books?id=reYnEQAAQBAJ>
- Umam, F., Satria, V. H., Sukri, H., & Adiputra, F. (2023). Machine Learning. Media Nusa Creative (MNC Publishing). <https://books.google.co.id/books?id=9RbCEQAAQBAJ>
- Yurika, C., Nanang, W., & dkk. (2026). Prediksi kelulusan mahasiswa tepat waktu menggunakan perbandingan algoritma random forest dan xgboost. *Jurnal Komputer dan Teknologi*, 4(1), 38–45.
- Zubair, A. (2026). Data Science Kompetitif dengan Kaggle: Langkah-Langkah Menuju Kemenangan. Penerbit Andi. https://books.google.co.id/books?id=OQ_IEQAAQBAJ