

EARLY WARNING SYSTEM KELULUSAN TEPAT WAKTU DENGAN MODEL XGBOOST DAN IDENTIFIKASI PEUBAH PENTING DENGAN MODEL LIME

IRSYAD PARAMA DWIPUTRA



**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH MATEMATIKA, SAINS DATA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “*Early Warning System* Kelulusan Tepat Waktu dengan Model XGBoost dan Identifikasi Peubah Penting dengan Model LIME” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dalam penyusunan karya ini penulis menggunakan bantuan kecerdasan buatan berupa Claude AI berbasis Claude Opus 4.8 untuk membantu penyuntingan bahasa, penyelarasan kalimat, dan peninjauan paragraf. Seluruh hasil penggunaan layanan tersebut telah ditinjau dan disunting kembali oleh penulis sesuai kebutuhan, sehingga penulis bertanggung jawab penuh atas isi karya tugas akhir ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2026

Irsyad Parama Dwiputra
G1401221101



ABSTRAK

IRSYAD PARAMA DWIPUTRA. *Early Warning System* Kelulusan Tepat Waktu dengan Model XGBoost dan Identifikasi Peubah Penting dengan Model LIME. Dibimbing oleh UTAMI DYAH SYAFITRI dan BAGUS SARTONO

Kelulusan mahasiswa tepat waktu merupakan indikator penting bagi kualitas dan efektivitas perguruan tinggi, sehingga mahasiswa berisiko perlu dikenali sejak dini untuk diberikan intervensi. Untuk itu, penelitian ini mengembangkan *early warning system* yang memprediksi kelulusan tepat waktu mahasiswa berdasarkan data akademik hingga semester lima dan semester enam di Sekolah Sains Data, Matematika, dan Informatika (SSMI), IPB University, termasuk pada angkatan yang belum menyelesaikan masa studi. Sistem ini dibangun dari data Angkatan 2021 yang memuat 19 peubah penjelas, mencakup prestasi akademik, konversi SKS kegiatan, dan latar belakang ekonomi. Jumlah mahasiswa yang lulus tepat waktu dan yang tidak tepat waktu cenderung tidak seimbang, untuk mengatasi ketidakseimbangan tersebut, data diseimbangkan menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE). Model klasifikasi dibangun menggunakan *Extreme Gradient Boosting* (XGBoost), yang performanya dioptimalkan melalui *hyperparameter tuning* dengan rancangan faktorial penuh dan dianalisis menggunakan ANOVA. Model akhir mencapai akurasi 78% dan F1-score terboboti 0,78 pada data uji. Untuk memahami faktor risiko pada tingkat individu, diterapkan *Local Interpretable Model-agnostic Explanations* (LIME) yang menghasilkan *feature importance* bagi setiap mahasiswa. Nilai kepentingan tersebut kemudian dikelompokkan menggunakan penggerombolan K-Means dan menghasilkan dua kelompok, yaitu kelompok berisiko dan tidak berisiko dimana pembeda utama antara keduanya adalah konversi SKS kegiatan dan latar belakang geografis. Berdasarkan klaster yang terbentuk diperlukan strategi yang berbeda disesuaikan dengan karakteristik tiap kelompok.

Kata kunci: *early warning system*, kelulusan tepat waktu, *K-Means clustering*, LIME, XGBoost



ABSTRACT

IRSYAD PARAMA DWIPUTRA. Early Warning System for Timely Graduation Using XGBoost and Identification of Important Features Using LIME. Supervised by UTAMI DYAH SYAFITRI and BAGUS SARTONO

Timely graduation is an important indicator of the quality and effectiveness of higher education institutions, so at-risk students must be identified early for intervention. This study develops an early warning system to predict timely graduation using academic data up to the fifth and sixth semesters at the School of Data Science, Mathematics, and Informatics (SSMI), IPB University, allowing application to cohorts still completing their studies. The 2021 cohort data used comprised 19 explanatory variables covering academic performance, extracurricular credit conversion, and economic background. Graduation Within the Standard Study Period (GSSP) students are usually unbalanced; the data were balanced using the Synthetic Minority Over-sampling Technique (SMOTE). The classification model was built using Extreme Gradient Boosting (XGBoost), whose performance was optimized through hyperparameter tuning with a full factorial design analyzed using ANOVA. The final model achieved an accuracy of 78% and a weighted F1-score of 0.78 on the test set. To understand risk factors at the individual level, Local Interpretable Model-agnostic Explanations (LIME) produced feature importance for each student. These values were then grouped using K-Means clustering into two groups, at-risk and not-at-risk, distinguished mainly by low extracurricular credit conversion and geographic background. This impacts strategies to speed up the study, as each group is different.

Keywords: *early warning system, K-Means clustering, LIME, timely graduation, XGBoost*



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

***EARLY WARNING SYSTEM* KELULUSAN TEPAT WAKTU DENGAN MODEL XGBOOST DAN IDENTIFIKASI PEUBAH PENTING DENGAN MODEL LIME**

IRSYAD PARAMA DWIPUTRA

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:
Dr. Ir. Indahwati M,Si.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Judul Skripsi : *Early Warning System* Kelulusan Tepat Waktu dengan Model XGBoost dan Identifikasi Peubah Penting dengan Model LIME

Nama : Irsyad Parama Dwiputra

NIM : G1401221101

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:

Dr. Utami Dyah Syafitri, S.Si., M.Si.

Pembimbing 2:

Dr. Bagus Sartono, S.Si, M.Si.

Diketahui oleh

Ketua Program Studi:

Dr. Bagus Sartono, S.Si, M.Si.

19780411 200501 1002

Tanggal Ujian:
31 Juli 2026

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Januari 2026 sampai bulan Agustus 2026 ini ialah *Early Warning System* Kelulusan Tepat Waktu , dengan judul “*Early Warning System* Kelulusan Tepat Waktu dengan Model XGBoost dan Identifikasi Peubah Penting dengan Model LIME”.

Penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah memberikan dukungan serta berperan dalam proses penulis menyelesaikan studi sarjana, antara lain ibu Dr. Utami Dyah Syafitri, S.Si., M.Si., dan Bapak Dr Bagus Sartono S.Si., M.Si., selaku dosen pembimbing yang sudah memberikan bimbingan, arahan, dan masukan yang sangat berarti dalam penyusunan karya ilmiah ini; Bapak Sapto Widyatmiko dan Ibu Umi Suryani sebagai orang tua penulis yang sudah memberikan dukungan dan support selama penulisan karya ilmiah; Ibu Dr. Ir. Indahwati M.Si. selaku dosen penguji luar komisi, Ibu Laily Nissa Atul Mualifah, S.Si., M.Si. selaku dosen moderator seminar hasil, Bapak Dr. Anwar Fitrianto, S.Si., M.Sc selaku dosen moderator kolokium yang sudah memberi saran dan masukan dalam karya ilmiah ini; Seluruh dosen dan tenaga pendidik Program Studi Statistika dan Sains Data yang telah memberi ilmu bermanfaat dan menunjang segala kebutuhan penulis selama perkuliahan dan penyusunan karya ilmiah ini; Indryani Sutisna sebagai pendukung terdepan penulis yang sudah memberi semangat dalam penulisan karya ilmiah ini; Sahabat dan teman-teman penulis, Excel, Ali, Aqil, Ryan, Fino, Perdana, Jidan, Farid Fahreza, Adel, Devva, Hani, Farid Leo, Diaz, Rakha, Sergei, dan Aping yang selalu memberi semangat dalam penulisan karya ilmiah ini; Marinestic Statistika 59 yang telah mendukung dan memberikan semangat kepada penulis dalam menyelesaikan tugas akhir.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan..

Bogor, Agustus 2026

Irsyad Parama Dwiputra



DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	x
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
TINJAUAN PUSTAKA	3
2.1 Kelulusan Mahasiswa Tepat Waktu	3
2.2 Transformasi Data	3
2.3 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	4
2.4 Rancangan Faktorial Penuh	4
2.5 <i>Extreme Gradient Boosting (XGBoost)</i>	5
2.6 Evaluasi Kinerja Model	6
2.7 <i>Local Interpretable Model-agnostic Explanations (LIME)</i>	7
2.8 <i>K-Means Clustering</i>	8
III METODE	11
3.1 Data	11
3.2 Prosedur Analisis Data	13
IV HASIL DAN PEMBAHASAN	16
4.1 Pra-pemrosesan Data	16
4.2 Eksplorasi Data	16
4.3 Partisi Data dan Penanganan Ketidakseimbangan Data	19
4.4 Pelatihan dan Evaluasi Model	19
4.5 Evaluasi Model XGBoost	20
4.6 Interpretasi Model dengan LIME	23
4.7 Penggerombolan K-Means Berbasis <i>Feature Importance</i> LIME	24
4.8 Prediksi Risiko Keterlambatan Angkatan 59	26
4.9 Prediksi Risiko Keterlambatan Angkatan 60	28
V SIMPULAN DAN SARAN	31
5.1 Simpulan	31
5.2 Saran	31
DAFTAR PUSTAKA	32
LAMPIRAN	34
RIWAYAT HIDUP	38



DAFTAR TABEL

1	Contoh representasi <i>one-hot encoding</i> untuk peubah tiga kategori	3
2	Struktur <i>confusion matrix</i> untuk klasifikasi dua kelas	6
3	Peubah yang digunakan dalam penelitian	11
4	Representasi <i>one-hot encoding</i> peubah nominal	13
5	<i>Hyperparameter</i> yang digunakan dalam penelitian	15
6	Hasil uji signifikansi <i>hyperparameter</i> (<i>three-way</i> ANOVA)	20
7	Kombinasi <i>hyperparameter</i> terbaik	20
8	Perbandingan kinerja model XGBoost pada data uji Angkatan 58	21
9	<i>Confusion matrix</i> model semester 1–6 pada data uji Angkatan 58	21
10	<i>Confusion matrix</i> model semester 1–5 pada data uji Angkatan 58	21
11	Perbandingan profil <i>cluster</i> K-Means berbasis <i>feature importance</i> LIME dua versi model Angkatan 58	24
12	Perbandingan distribusi prediksi risiko kelulusan Angkatan 59 dua versi model	26
13	Kategori prediksi risiko kelulusan tepat waktu Angkatan 60	28

DAFTAR GAMBAR

1	Ilustrasi skema <i>5-fold cross validation</i>	6
2	Ilustrasi pengelompokan K-Means (a) dan penentuan jumlah cluster optimal melalui silhouette score (b)	10
3	Diagram alur tahapan penelitian	13
4	Distribusi peubah respon kelulusan tepat waktu mahasiswa Angkatan 58	16
5	Proporsi mahasiswa lulus tepat waktu pada setiap program studi	17
6	Persentase kelulusan mahasiswa berdasarkan jumlah mata kuliah nilai D	17
7	Persentase kelulusan mahasiswa berdasarkan jumlah mata kuliah diulang	18
8	Rata-rata IP per semester pada kelompok lulus tepat waktu dan tidak tepat waktu	18
9	Proporsi dan jumlah mahasiswa lulus tepat waktu berdasarkan kategori SKSKonversiEC	19
10	Tingkat kepentingan fitur model semester 1–6 (a) dan model semester 1–5 (b)	22
11	Interpretasi LIME tiga contoh mahasiswa berisiko: (a) mahasiswa pertama (peluang 2,9%), (b) mahasiswa kedua (peluang 19,3%), dan (c) mahasiswa ketiga (peluang 10,7%)	23
12	Profil <i>cluster</i> K-Means berbasis <i>feature importance</i> LIME model semester 1–6 Angkatan 58	25
13	Profil <i>cluster</i> K-Means berbasis <i>feature importance</i> LIME model semester 1–5 Angkatan 58	26
14	Profil <i>cluster</i> K-Means berbasis <i>feature importance</i> LIME model semester 1–6 Angkatan 59	27
15	Profil <i>cluster</i> K-Means berbasis <i>feature importance</i> LIME model semester 1–5 Angkatan 59	27
16	Profil <i>cluster</i> K-Means berbasis <i>feature importance</i> LIME model semester 1–6 (dengan imputasi) Angkatan 60	29
17	Profil <i>cluster</i> K-Means berbasis <i>feature importance</i> LIME model semester 1–5 Angkatan 60	29



DAFTAR LAMPIRAN

1	Lampiran 1 Tabel ANOVA lengkap model semester 1–6	35
2	Lampiran 2 Tabel ANOVA lengkap model semester 1–5	35
3	Lampiran 3 Proporsi mahasiswa lulus tepat waktu berdasarkan kategori SKS kumulatif semester 6	36
4	Lampiran 4 Proporsi mahasiswa lulus tepat waktu berdasarkan jalur masuk	36
5	Lampiran 5 Proporsi mahasiswa lulus tepat waktu berdasarkan status kepemilikan KIP	36
6	Lampiran 6 Proporsi mahasiswa lulus tepat waktu berdasarkan golongan UKT	37
7	Lampiran 7 Proporsi mahasiswa lulus tepat waktu berdasarkan jenis kelamin	37
8	Lampiran 8 Proporsi mahasiswa lulus tepat waktu berdasarkan asal daerah	37