

IDENTIFIKASI PEUBAH PENTING DALAM KLASIFIKASI INDEKS PRESTASI MAHASISWA SNBP IPB UNIVERSITY ANGKATAN 2024 MENGGUNAKAN XGBOOST

SYIFA KHAIRUNNISA



**PROGRAM STUDI STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**

- 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 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.

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Identifikasi Peubah Penting dalam Klasifikasi Indeks Prestasi Mahasiswa SNBP IPB University Angkatan 2024 Menggunakan *XGBoost*” 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.

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

Bogor, Juni 2025

Syifa Khairunnisa
G1401211012

- 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.



ABSTRAK

SYIFA KHAIRUNNISA. Identifikasi Peubah Penting dalam Klasifikasi Indeks Prestasi Mahasiswa SNBP IPB University Angkatan 2024 Menggunakan *XGBoost*. Dibimbing oleh UTAMI DYAH SYAFITRI dan LAILY NISSA ATUL MUALIFAH.

Perguruan tinggi memegang kunci dalam menghasilkan sumber daya manusia yang berkualifikasi tinggi. Untuk mendukung hal tersebut, pemerintah menyediakan beberapa jalur masuk ke perguruan tinggi, salah satunya SNBP (Seleksi Nasional Berdasarkan Prestasi). SNBP membuka kesempatan untuk para siswa yang berprestasi di bangku sekolah untuk melanjutkan pendidikan ke jenjang lebih tinggi. Di perguruan tinggi, Indeks Prestasi (IP) menjadi indikator utama dalam mengukur keberhasilan akademik mahasiswa. Untuk memahami faktor-faktor yang memengaruhi capaian IP mahasiswa SNBP IPB University, studi ini menerapkan metode klasifikasi *XGBoost*. Salah satu keunggulan *XGBoost* adalah mampu mengidentifikasi peubah penting yang berpengaruh terhadap hasil akademik. Data yang digunakan pada penelitian ini merupakan data IP mahasiswa jalur SNBP 2024 pada semester pertama yang $\geq 2,10$ dengan 14 peubah penjelas. Data yang digunakan tidak seimbang, sehingga SMOTE diaplikasikan untuk menanganinya. Selain itu, *Fractional Factorial Design* diterapkan untuk menentukan *hyperparameter* yang signifikan terhadap kinerja model *XGBoost*. Model terbaik yang dipilih adalah model klasifikasi 3 kelas tanpa penanganan ketidakseimbangan data dengan akurasi sebesar 63,85%. *F1-score* sebesar 59,64% menunjukkan bahwa model mampu mengklasifikasikan data lebih seimbang antara ketepatan prediksi dan kemampuan menangkap kelas yang relevan. Adapun peubah penting dalam klasifikasi IP adalah pilihan program studi mahasiswa, prestasi akademik, gaya belajar, dan program studi tempat mahasiswa diterima.

Kata kunci: indeks prestasi mahasiswa, klasifikasi, peubah penting, SMOTE, *XGBoost*.

ABSTRACT

SYIFA KHAIRUNNISA. Identification of Feature Important in the Classification of Student Achievement Index of SNBP IPB University Class of 2024 Using XGBoost. Supervised by UTAMI DYAH SYAFITRI and LAILY NISSA ATUL MUALIFAH.

Universities hold a key in producing highly qualified human resources. To support that, the government provides one of the university entrances, SNBP (National Achievement-Based Selection). SNBP opens opportunities for students with achievement in high school to continue their education to a higher level. In University, Grade Point Average (GPA) is the main indicator for student success assessment. To analyze the factors that affect GPA of IPB University SNBP students, this study applied the XGBoost classification method. One of the advantages of XGBoost is it's ability to identify features that have the most effect on GPA. The data used in this study is GPA of SNBP 2024 students in first semester which is $\geq 2,10$ with 14 explanatory variables. The data used is imbalanced, so SMOTE is applied to handle it. In addition, Fractional Factorial Design is applied to determine hyperparameters that significant to the XGBoost model's performance. The selected XGBoost model is a 3-class classification model without SMOTE with an accuracy level of 63,85%. F1-score level of 59,64% indicates that this model can classify data more balanced between prediction exactness and the ability to capture relevant classes. The important features in GPA's classification were the first-choice study program, academic achievement, learning style, and study program where student was accepted.

Keywords: classification, feature importance, grade point average, SMOTE, 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 2025
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.

IDENTIFIKASI PEUBAH PENTING DALAM KLASIFIKASI INDEKS PRESTASI MAHASISWA SNBP IPB UNIVERSITY ANGKATAN 2024 MENGGUNAKAN XGBOOST

SYIFA KHAIRUNNISA

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

**PROGRAM STUDI STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@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 : Identifikasi Peubah Penting dalam Klasifikasi Indeks Prestasi
Mahasiswa SNBP IPB University Angkatan 2024 Menggunakan
XGBoost

Nama : Syifa Khairunnisa
NIM : G1401211012

@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.

Disetujui oleh

Pembimbing 1:
Dr. Utami Dyah Syafitri, S.Si, M.Si.



Pembimbing 2:
Laily Nissa Atul Mualifah, M.Si.

Diketahui oleh

Ketua Program Studi:
Dr. Bagus Sartono, S.Si, M.Si.
NIP. 19780411 200501 1002



Tanggal Ujian:
10 Juni 2025

Tanggal Lulus:
(tanggal penandatanganan oleh Dekan
Fakultas/Sekolah ...)



@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.

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 2025 sampai bulan Juni 2025 ini ialah Indeks Prestasi Mahasiswa, dengan judul “Identifikasi Peubah Penting dalam Klasifikasi Indeks Prestasi Mahasiswa SNBP IPB University Angkatan 2024 Menggunakan *XGBoost*”.

Terselesaikannya karya ilmiah ini tidak terlepas dari dukungan dan peran dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Atjep Saktiawan, Ibu Dwi Irawati, Kakak Fitri Azzahrah, dan Adik Shabina Nurul Zakiyyah selaku kedua orang tua dan keluarga yang selalu memberikan dukungan, doa, dan kasih sayang yang tiada henti.
2. Ibu Dr. Utami Dyah Syafitri, S.Si, M.Si. dan Ibu Laily Nissa Atul Muallifah, M.Si. selaku dosen pembimbing yang telah membimbing, mengarahkan, membantu, memotivasi, dan memberikan saran serta masukan sehingga karya ilmiah ini dapat terselesaikan dengan baik.
3. Bapak Dr. Septian Rahardiantoro, S.Stat., M.Si. selaku penguji luar komisi, Ibu Rahma Anisa, S.Stat., M.Si. selaku moderator seminar hasil, dan Bapak La Ode Abdul Rahmad, S.Si., M.Si. selaku moderator kolokium atas saran dan masukan untuk penulisan karya ilmiah ini. Tidak lupa, Bapak Dr. Agus Mohamad Soleh, S.Si., M.T. yang telah bersedia memberikan pandangan serta saran.
4. Dosen dan tenaga pendidik Statistika dan Sains Data IPB University yang telah memberikan ilmu yang bermanfaat dan bantuan dalam kelancaran administrasi selama masa perkuliahan dan penyusunan karya ilmiah ini.
5. Direktorat Administrasi Pendidikan dan Penerimaan Mahasiswa Baru IPB University yang telah memberikan penulis kesempatan untuk menggunakan data mahasiswa dalam penelitian ini.
6. Davina Rachmadyanti, Salsabila Dwi Rahmi, Gladys Adya Zafira, dan Cahyani Dyah Rofiana selaku sahabat terdekat penulis yang selalu menemani, membantu, dan memberikan dukungan selama masa kuliah dan penyusunan karya ilmiah ini.
7. Kheni Hikmah Lestari, Arfiah Kania Sektiaruni, Alfikri Ihsan, Zulfa Hafizhoh, Muhamad Fawaz Zidan, Akmal Basis Jatining Kusumah, Oktavia Galih Pratiwi, Fajryanti Kusuma Wardani, Yasmin Azimah Wafa, Natasha Muti Hafiza, Ravi Mahesa Pramudya, Raziqizzan Putrandi, Dhiya Khalishah TS, dan teman-teman Statistika 58 lain yang telah menemani dan memberikan dorongan yang positif bagi penulis selama masa kuliah dan penyusunan karya ilmiah ini.
8. Raidha Hasna Kusuma, Anissa Alivia Putri, Rifa Ramadhani, Syavina Arsila Jannah, Erlyn Septiyana Putri, Maeyza Ardita Arini, Desi Nur Fitriani, Ervina Kurniasih Wibowo, Amelia Widasari, dan teman-teman penulis lainnya yang telah mendengarkan dan memberi semangat selama penyusunan karya ilmiah ini.



9. Diri penulis sendiri yang telah berjuang sejauh ini.
10. Seluruh pihak terkait yang telah membantu dalam penyusunan karya ilmiah ini yang tidak dapat disebutkan satu persatu.

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

Bogor, Juni 2025

Syifa Khairunnisa

DAFTAR ISI

DAFTAR TABEL	xii
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	3
II TINJAUAN PUSTAKA	4
2.1 Indeks Prestasi (IP) Mahasiswa	4
2.2 <i>Fractional Factorial Design (FFD)</i>	4
2.3 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	6
2.4 <i>eXtreme Gradient Boosting (XGBoost) Classifier</i>	7
2.5 Evaluasi Kebaikan Model	10
2.6 <i>Feature Importance</i>	11
III METODE	13
3.1 Data	13
3.2 Prosedur Analisis Data	14
IV HASIL DAN PEMBAHASAN	19
4.1 Praproses Data	19
4.2 Eksplorasi Data	20
4.3 <i>Hyperparameter Tuning</i>	24
4.4 Pembagian Data dan Penanganan Ketidakseimbangan Data	26
4.5 Pelatihan Model Klasifikasi dan Evaluasi	28
4.6 Tingkat Kepentingan Peubah	30
V SIMPULAN DAN SARAN	34
5.1 Simpulan	34
5.2 Saran	34
DAFTAR PUSTAKA	35
LAMPIRAN	38
RIWAYAT HIDUP	44



DAFTAR TABEL

1	Matriks konfusi klasifikasi biner	10
2	Matriks konfusi klasifikasi 3 kelas	10
3	Formulasi metrik evaluasi	11
4	Peubah yang digunakan	13
5	Kategorisasi kelas IP semester 1 skema 1 (3 kelas)	15
6	Kategorisasi kelas IP semester 1 skema 2 (4 kelas)	15
7	Nilai <i>hyperparameter</i> yang dicobakan	16
8	Peubah baru yang dibentuk	19
9	Hasil signifikansi <i>hyperparameter</i>	25
10	Kombinasi nilai <i>hyperparameter</i> terbaik	26
11	Persentase peubah <i>Y</i> pada data latih dan uji	26
12	Komposisi data latih sebelum dan sesudah SMOTE	27
13	Contoh data pada peubah yang digunakan	27
14	Contoh data SMOTE pada peubah yang digunakan	28
15	Rataan metrik evaluasi	29
16	<i>Confusion matrix</i> hasil prediksi dari Model 1	30
17	Karakteristik peubah penting masing-masing kelas	32

DAFTAR GAMBAR

1	Contoh grafik desain fractional factorial 3^{7-3} dengan $I = ABCE = ABDF = AB^2C^2DG$	5
2	Ilustrasi pembangkitan data sintesis dengan SMOTE	7
3	Ilustrasi alur kerja <i>XGBoost</i>	9
4	Diagram alir analisis data pada penelitian	18
5	Diagram kue kurikulum sekolah dan jurusan SMA	20
6	Histogram sebaran IP semester 1 mahasiswa S1 SNBP angkatan 2024	21
7	Sebaran IP semester 1 berdasarkan fakultas diterima	21
8	Sebaran IP semester 1 berdasarkan kode program studi diterima	22
9	Diagram batang tipe belajar (a) sebaran IP semester 1 berdasarkan tipe belajar mahasiswa (b)	23
10	Proporsi amatan masing-masing kelas pada skema 1 (a) pada skema 2 (b)	24
11	Pohon klasifikasi Model 1 iterasi keempat	30
12	<i>Feature importance</i> dalam klasifikasi IP Semester 1 berdasarkan nilai <i>gain</i>	31

DAFTAR LAMPIRAN

1	Lampiran 1 Keterangan Kode Program Studi	39
2	Lampiran 2 Keterangan Kategori Buku Perak	40
3	Lampiran 3 Keterangan Klaster Sekolah	40

4	Lampiran 4 Tabel ANOVA untuk skema 1	40
5	Lampiran 5 Tabel ANOVA untuk skema 2	41
6	Lampiran 6 <i>Confusion matrix</i> Model 2	41
7	Lampiran 7 <i>Confusion matrix</i> Model 3	41
8	Lampiran 8 <i>Confusion matrix</i> Model 4	41
9	Lampiran 9 Nilai <i>gain</i> dari peubah yang digunakan	42
10	Lampiran 10 Pohon pertama pada iterasi pertama	43

@Hak cipta milik IPB University

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.



@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.