



IDENTIFIKASI PEUBAH BERPENGARUH TERHADAP INDEKS PRESTASI MAHASISWA SNBP DIPLOMA IV IPB UNIVERSITY TAHUN 2024 DENGAN METODE *RANDOM FOREST* DAN *XGBOOST*

MUHAMAD FAWAZ ZIDAN



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.

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Identifikasi Peubah Berpengaruh terhadap Indeks Prestasi Mahasiswa SNBP Diploma IV IPB University Tahun 2024 dengan Metode *Random Forest* dan *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

Muhamad Fawaz Zidan
G1401211051



ABSTRAK

MUHAMAD FAWAZ ZIDAN. Identifikasi Peubah Berpengaruh terhadap Indeks Prestasi Mahasiswa SNBP Diploma IV IPB University Tahun 2024 dengan Metode *Random Forest* dan *XGBoost*. Dibimbing oleh UTAMI DYAH SYAFITRI dan ERFIANI.

Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang memengaruhi Indeks Prestasi (IP) mahasiswa tahun pertama program D4 IPB University angkatan 2024. Mahasiswa ini masuk melalui berbagai jalur penerimaan yang diatur dalam Peraturan Menteri Pendidikan Nomor 48 Tahun 2022. Setelah menyelesaikan semester pertama, mahasiswa memperoleh IP yang menjadi dasar penentuan beban studi pada semester-semester berikutnya. Penelitian sebelumnya menunjukkan bahwa latar belakang sekolah dan jalur penerimaan berpengaruh signifikan terhadap performa akademik. Penelitian ini menerapkan metode *ensemble learning*, yaitu *Random Forest* dan *Extreme Gradient Boosting (XGBoost)*, yang dapat meningkatkan akurasi prediksi dan meminimalkan *overfitting*. Data diperoleh dari Direktorat Administrasi Pendidikan dan Penerimaan Mahasiswa Baru (DAPPMB) IPB tahun 2024. Data diproses menggunakan teknik klasifikasi dengan *cross-validation*, serta dievaluasi menggunakan metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-Score*. Tiap metode dimodelkan dengan dan tanpa menggunakan *class weighting* serta *hyperparameter* untuk hasil yang lebih optimal. *F1-Score* model terbaik dari metode *Random Forest* adalah 0.644 yang diperoleh tanpa *class weighting*, sedangkan *XGBoost* adalah 0.431 yang diperoleh tanpa *class weighting*. Hasil penelitian menunjukkan bahwa model *Random Forest* tanpa *class weighting* memberikan performa terbaik dibandingkan model lainnya, dengan *F1-Score* tertinggi sebesar 0.644. Peubah berpengaruh dipilih dari model terbaik, yaitu *Random Forest* tanpa *class weighting*. Peubah yang berpengaruh signifikan terhadap indeks prestasi mahasiswa D4 IPB University tahun 2024, antara lain rata-rata total rapor, rata-rata syarat rapor, dan program studi diterima.

Kata kunci: *ensemble learning*, indeks prestasi, *random forest*, *xgboost*, peubah berpengaruh

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan 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 mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

ABSTRACT

MUHAMAD FAWAZ ZIDAN. Identification of Influential Variables on the Grade Point Average of SNBP Diploma IV Students at IPB University in 2024 Using Random Forest and XGBoost Methods. Supervised by UTAMI DYAH SYAFITRI and ERFIANI.

This study aims to identify the factors influencing the Grade Point Average (GPA) of first-year D4 students at IPB University in the 2024 cohort. These students admitted through various admission pathways regulated under the Ministry of Education Regulation No. 48 of 2022. After completing their first semester, students receive a GPA that is the basis for determining their academic load in subsequent semesters. Previous studies have shown that school background and admission route significantly affect academic performance. This research applied ensemble learning methods, namely Random Forest and Extreme Gradient Boosting (XGBoost), known for enhancing prediction accuracy and minimizing overfitting. The data were obtained from the Directorate of Academic Administration and New Student Admission (DAPPMB) of IPB in 2024. The data were processed using classification techniques with cross-validation and evaluated using performance metrics such as accuracy, precision, recall, and F1-Score. Each method was modeled both with and without class weighting and hyperparameter tuning to achieve more optimal results. The best F1-Score from the Random Forest method was 0.644, obtained without class weighting, while XGBoost achieved its best F1-Score of 0.431, also without class weighting. The results indicate that the Random Forest model without class weighting provided the best performance among all models, with the highest F1-Score of 0.644. Influential variables were selected based on this best-performing model, that is Random Forest without class weighting. The variables that significantly influenced the GPA of D4 students at IPB University in 2024 include the average total report card score, average requirement report card score, and accepted study program.

Keywords: ensemble learning, grade point, random forest, xgboost, feature importance



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 2025¹
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 BERPENGARUH TERHADAP INDEKS PRESTASI MAHASISWA SNBP DIPLOMA IV IPB UNIVERSITY TAHUN 2024 DENGAN METODE *RANDOM FOREST* DAN *XGBOOST*

MUHAMAD FAWAZ ZIDAN

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.



@Hak cipta milik IPB University

Judul Skripsi : Identifikasi Peubah Berpengaruh terhadap Indeks Prestasi
Mahasiswa SNBP Diploma IV IPB University Tahun 2024
dengan Metode *Random Forest* dan *XGBoost*

Nama : Muhamad Fawaz Zidan
NIM : G1401211051

Disetujui oleh

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



Pembimbing 2:
Dr. Ir. Erfiani, M.Si.



Diketahui oleh

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



Tanggal Ujian:
(5 Juni 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.





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 Februari 2025 sampai bulan Juni 2025 ini ialah Indeks Prestasi (IP) Mahasiswa, dengan judul “Identifikasi Peubah Berpengaruh terhadap Indeks Prestasi Mahasiswa SNBP Diploma IV IPB University Tahun 2024 dengan Metode *Random Forest* dan *XGBoost*”.

Terselesainya penulisan ini tidak lepas dari dukungan dan bantuan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Ahmad Fuadi dan Ibu Wahyu Tri Kencana, orang tua yang selalu percaya dengan seluruh mimpi putra tunggalnya. Terima kasih atas doa dalam diam, peluh yang tak pernah ditagih balas, dan cinta yang tak mengenal jeda, serta ketulusan hati yang jadi pondasi penulis hingga titik ini;
2. Ibu Dr. Utami Dyah Syafitri, S.Si., M.Si. dan Ibu Dr. Ir. Erfiani, M.Si. selaku dosen pembimbing. Terima kasih atas bimbingan yang sabar, arahan yang tajam namun membangun, serta kesediaan meluangkan waktu di tengah kesibukan. Tidak lupa, Bapak Dr. Agus Mohamad Soleh, S.Si., M.T. yang telah membantu memberikan pandangan dan masukan solutif terkait penelitian ini;
3. Bapak Dr. Ir. Budi Susetyo, M.S. selaku dosen penguji pada sidang skripsi. Terima kasih atas saran yang membangun serta pertanyaan kritis agar penulisan ini lebih baik dan bermanfaat kedepannya;
4. Dosen dan tenaga pendidik Program Studi Statistika dan Sains Data IPB University yang telah memberikan ilmu, arahan, serta bantuan, baik dalam proses akademik maupun administrasi;
5. Direktorat Administrasi Pendidikan IPB University yang telah memberikan serta mengizinkan penulis untuk menggunakan data mahasiswa dalam kepentingan penulisan karya ilmiah ini;
6. Teman-teman Statistika 58 yang telah kebersamai penulis dalam menyusun karya ilmiah ini;
7. Sahabat-sahabat penulis, Pratama, Adeva, Reynd, Raziq, Hilmi, Akmal, Fadly, Rifqi, Rafee, Ravi, Alfikri, Syifa, Cahyo, dan Luthfi. Terima kasih telah menjadi pelabuhan tawa di tengah lelah, tempat berbagi gelisah tanpa takut dihakimi, dan pengingat bahwa perjalanan ini tidak pernah harus dijalani sendiri;
8. Kepada perempuan dengan hadir yang menenangkan. Terima kasih atas pengertian tanpa banyak tapi, kehadiran tanpa banyak bising, dan dukungan tanpa letih hingga saat ini. Semoga waktu terus berpihak dengan membawa kita ke arah yang satu;
9. Kepada diri ini, penulis sekaligus putra tunggal dengan ribuan mimpi. Terima kasih telah tetap melangkah dalam diam, menata kembali harapan yang hampir pupus, dan menjaga nyala meski kadang redup. Semoga, segala luka, ragu, lelah, dan harapan, yang dipeluk dalam diam hingga saat ini, perlahan menjadi pelajaran untuk tetap tumbuh dan melangkah.

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

Bogor, Juni 2025

Muhamad Fawaz Zidan



@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.

DAFTAR ISI

DAFTAR TABEL	ix
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	ix
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Manfaat	2
II TINJAUAN PUSTAKA	3
2.1 Indeks Prestasi (IP)	3
2.2 <i>Class Weighting</i>	3
2.3 <i>K-Fold Cross-Validation (K-Fold CV)</i>	4
2.4 <i>Random Forest</i>	4
2.5 <i>Extreme Gradient Boosting (XGBoost)</i>	5
2.6 Metrik Evaluasi	6
2.7 <i>Feature Importance</i>	7
III METODE	9
3.1 Data	9
3.2 Prosedur Analisis Data	10
IV HASIL DAN PEMBAHASAN	13
4.1 Praproses Data	13
4.2 Eksplorasi Data	13
4.3 Partisi Data dan Penanganan <i>Unbalanced Data</i>	15
4.4 Pelatihan dan Evaluasi Model	16
4.5 Identifikasi Peubah Berpengaruh	20
V SIMPULAN DAN SARAN	21
5.1 Simpulan	21
5.2 Saran	21
DAFTAR PUSTAKA	22
LAMPIRAN	24
RIWAYAT HIDUP	28

DAFTAR TABEL

1	<i>Confussion matrix</i> untuk untuk hasil prediksi data uji masing-masing kelas	7
2	Peubah respons dan peubah penjelas yang digunakan pada penelitian	9
3	Kategorisasi Indeks Prestasi (IP) mahasiswa	10
4	<i>Hyperparameter</i> yang digunakan	10
5	Hasil evaluasi masing-masing model	17
6	<i>Confussion matrix</i> model <i>Random Forest</i> dengan <i>class weighting</i>	17
7	<i>Confussion matrix</i> model <i>Random Forest</i> tanpa <i>class weighting</i>	17
8	<i>Confussion matrix</i> model <i>XGBoost</i> dengan <i>class weighting</i>	18
9	<i>Confussion matrix</i> model <i>XGBoost</i> tanpa <i>class weighting</i>	18
10	Mekanisme internal model <i>Random Forest</i> pada sampel data uji	19
11	Rekapitulasi hasil <i>voting Random Forest</i> pada sampel data uji	19

DAFTAR GAMBAR

1	Ilustrasi pembagian <i>fold</i> (bagian) pada K-Fold CV	4
2	Diagram alir analisis data penelitian	12
3	Indeks prestasi mahasiswa SNBP D4 IPB University	13
4	Indeks prestasi mahasiswa SNBP pada tiap program studi D4 IPB University	14
5	Hasil <i>Talent Mapping</i> (TM) mahasiswa SNBP D4 IPB University	14
6	Jurusan sekolah asal mahasiswa SNBP D4 IPB University	15
7	Distribusi kelas pada data sebelum dan sesudah penanganan ketidakseimbangan	16
8	Ilustrasi alur prediksi pada sampel pohon <i>Random Forest</i>	19
9	Peubah yang berpengaruh terhadap hasil prediksi	20

DAFTAR LAMPIRAN

1	Metrik evaluasi model dari metode <i>Random Forest</i> tanpa <i>class weighting</i>	25
2	Metrik evaluasi model dari metode <i>Random Forest</i> dengan <i>class weighting</i>	25
3	Persentase peubah berpengaruh terhadap indeks prestasi mahasiswa	25
4	Karakteristik gerombol klaster sekolah	25
5	Buku perak	25
6	Program studi vokasi (D4) IPB University	26
7	Kategori hasil <i>talent mapping</i>	25