

IDENTIFIKASI PEUBAH PENTING TERHADAP INDEKS PRESTASI MAHASISWA D-IV JALUR SNBP DENGAN *RANDOM FOREST* DAN *EASYENSEMBLE*

ELKE FRIDA RAHMAWATI



**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.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Identifikasi Peubah Penting terhadap Indeks Prestasi Mahasiswa D-IV Jalur SNBP dengan *Random Forest* dan *EasyEnsemble*” 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 tugas akhir ini, penulis memanfaatkan layanan kecerdasan buatan, yaitu ChatGPT dan Gemini, sebagai alat bantu dalam penyuntingan bahasa, penyelarasan struktur kalimat, serta peninjauan kembali isi paragraf. Seluruh hasil yang diperoleh dari penggunaan layanan tersebut telah ditelaah dan disunting kembali oleh penulis sesuai dengan kaidah penulisan ilmiah. Dengan demikian, penulis bertanggung jawab sepenuhnya atas isi dan substansi tugas akhir ini.

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

Bogor, Agustus 2026

Elke Frida Rahmawati
G1401221025



ABSTRAK

ELKE FRIDA RAHMAWATI. Identifikasi Peubah Penting terhadap Indeks Prestasi Mahasiswa D-IV Jalur SNBP dengan *Random Forest* dan *EasyEnsemble*. Dibimbing oleh UTAMI DYAH SYAFITRI dan LAILY NISSA ATUL MUALIFAH.

Indeks Prestasi (IP) Semester 1 merupakan indikator keberhasilan akademik awal mahasiswa yang dipengaruhi oleh berbagai faktor. Sementara itu, distribusi kategori IP pada mahasiswa D-IV jalur Seleksi Nasional Berdasarkan Prestasi (SNBP) cenderung tidak seimbang sehingga dapat menurunkan kemampuan model klasifikasi dalam mengenali kelas minoritas. Penelitian ini bertujuan membandingkan performa *Random Forest* standar dan *Random Forest* berbasis *EasyEnsemble* dalam mengklasifikasikan kategori IP Semester 1 serta mengidentifikasi peubah yang berkontribusi terhadap hasil klasifikasi menggunakan *SHapley Additive exPlanations* (SHAP). Data yang digunakan merupakan data primer mahasiswa D-IV angkatan 62 jalur SNBP IPB University tahun 2025. Evaluasi model dilakukan menggunakan *repeated stratified hold out* sebanyak lima kali ulangan dengan kombinasi *hyperparameter ntree* (50, 100, 150) dan *mtry* (3, 5, 7). Performa model dibandingkan menggunakan *Balanced Accuracy*, *Macro F1-Score*, dan *Geometric Mean (G-Mean)*, sedangkan pengaruh *hyperparameter* dianalisis menggunakan ANOVA faktorial dua arah. Hasil penelitian menunjukkan bahwa *Random Forest* berbasis *EasyEnsemble* pada skema tiga kelas menghasilkan performa terbaik dengan peningkatan *Balanced Accuracy* dan *G-Mean* dibandingkan *Random Forest* standar. Analisis SHAP menunjukkan bahwa Rataan Pendidikan Kewarganegaraan, Rataan Bahasa Indonesia, Rataan Matematika, Rataan Prasyarat, dan Rataan Total merupakan peubah yang paling berkontribusi terhadap klasifikasi kategori IP Semester 1.

Kata kunci: *easyensemble*, indeks prestasi, *random forest*, SHAP, SNBP

ABSTRACT

ELKE FRIDA RAHMAWATI. Identification of Important Variables Influencing the Grade Point Average of D-IV Students Admitted Through the SNBP Pathway Using Random Forest and EasyEnsemble. Supervised by UTAMI DYAH SYAFITRI and LAILY NISSA ATUL MUALIFAH.

First-semester Grade Point Average (GPA) is an important indicator of students' early academic performance. However, the distribution of GPA categories among D-IV students admitted through the National Selection Based on Achievement (SNBP) pathway is imbalanced, which may reduce the ability of classification models to identify minority classes. This study compared the performance of the standard Random Forest and EasyEnsemble-based Random Forest models for classifying first-semester GPA categories and identified important variables contributing to the classification using SHapley Additive exPlanations (SHAP). The study used primary data from D-IV students of the 62nd cohort admitted through the SNBP pathway at IPB University in 2025. Model evaluation was conducted using repeated stratified holdout with five repetitions and combinations of *n*tree (50, 100, and 150) and *m*try (3, 5, and 7). Model performance was assessed using Balanced Accuracy, Macro F1-Score, and Geometric Mean (G-Mean), while the effects of hyperparameters were evaluated using two-way ANOVA. The results showed that the EasyEnsemble-based Random Forest with the three-class scheme achieved the best performance by improving Balanced Accuracy and G-Mean compared with the standard Random Forest. SHAP analysis indicated that Civic Education average score, Indonesian Language average score, Mathematics average score, prerequisite subject average score, and overall report card average score were the most influential variables in classifying first-semester GPA categories.

Keywords: easyensemble, grade point average, random forest, SHAP, SNBP.

IDENTIFIKASI PEUBAH PENTING TERHADAP INDEKS PRESTASI MAHASISWA D-IV JALUR SNBP DENGAN *RANDOM FOREST* DAN *EASYENSEMBLE*

ELKE FRIDA RAHMAWATI

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:

1 Dr. Anwar Fitrianto, S.Si. M.Sc



@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 terhadap Indeks Prestasi Mahasiswa
D-IV Jalur SNBP dengan *Random Forest* dan *EasyEnsemble*

Nama : Elke Frida Rahmawati

NIM : G1401221025

Disetujui oleh

Pembimbing 1:

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



Pembimbing 2:

Laily Nissa Atul Kualifah, S.Si, M.Si.

Diketahui oleh

Ketua Program Studi:

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

NIP. 19780411 200501 1002

Tanggal Ujian:
(4 Agustus 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 Februari 2026 sampai bulan Juni 2026 ini ialah Indeks Prestasi Mahasiswa, dengan judul “Identifikasi Peubah Penting terhadap Indeks Prestasi Mahasiswa D-IV Jalur SNBP dengan *Random Forest* dan *EasyEnsemble*”.

Penyusunan skripsi ini tidak terlepas dari dukungan, bimbingan, doa, dan bantuan dari berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada Bapak Kariyana, Ibu Siti Masikah, dan Elfa Abidah Salsabila selaku keluarga yang senantiasa memberikan doa, kasih sayang, dukungan, dan kepercayaan kepada penulis selama menempuh pendidikan. Penulis juga menyampaikan terima kasih kepada Ibu Dr. Utami Dyah Syafitri, S.Si., M.Si. dan Ibu Laily Nissa Atul Mualifah, S.Si., M.Si. selaku dosen pembimbing atas waktu, perhatian, kesabaran, bimbingan, motivasi, serta arahan yang diberikan selama proses penyusunan skripsi ini. Ucapan terima kasih juga penulis sampaikan kepada Bapak Dr. Anwar Fitrianto, S.Si., M.Sc. selaku moderator kolokium dan Bapak Dr. Bagus Sartono, S.Si., M.Si. selaku moderator seminar hasil atas saran dan masukan yang diberikan. Penulis juga menyampaikan penghargaan kepada seluruh dosen dan tenaga kependidikan Program Studi Statistika dan Sains Data IPB University atas ilmu, bantuan, dan pelayanan akademik selama masa perkuliahan, serta kepada Direktorat Administrasi Pendidikan dan Penerimaan Mahasiswa Baru IPB University yang telah memberikan izin penggunaan data mahasiswa dalam penelitian ini. Ucapan terima kasih turut penulis sampaikan kepada sahabat-sahabat penulis, teman-teman satu bimbingan, keluarga Statistika 59 (Marinestic), serta sahabat-sahabat sejak SMA yang senantiasa memberikan doa, semangat, dukungan, dan kebersamaan selama proses penyusunan skripsi ini. Terakhir, penulis mengucapkan terima kasih kepada seluruh pihak yang telah membantu dan memberikan dukungan selama penyusunan skripsi ini yang tidak dapat disebutkan satu per satu.

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

Bogor, Agustus 2026

Elke Frida Rahmawati



DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Manfaat	2
II TINJAUAN PUSTAKA	4
2.1 Indeks Prestasi	4
2.2 <i>Repeated Stratified Hold Out</i>	4
2.3 <i>EasyEnsemble</i>	5
2.4 <i>Random Forest</i>	6
2.5 Analisis Varians (<i>Analysis of Variance</i>)	7
2.6 Evaluasi Keباikan Model	8
2.7 SHAP (<i>SHapley Additive exPlanations</i>)	10
III METODE	12
3.1 Data	12
3.2 Metode Penelitian	12
IV HASIL DAN PEMBAHASAN	15
4.1 Praproses Data	15
4.2 Eksplorasi Data	15
4.3 Pembagian Data	18
4.4 Penanganan Ketidakseimbangan Data dengan <i>EasyEnsemble</i>	19
4.5 Pengujian <i>Hyperparameter</i> dengan ANOVA Faktorial Dua Arah pada Model <i>Random Forest</i>	20
4.6 Perbandingan Model <i>Random Forest</i> Standar dengan <i>Random Forest EasyEnsemble</i>	22
4.7 Interpretasi Model <i>Random Forest EasyEnsemble</i> menggunakan Pohon Keputusan Tunggal	26
4.8 Identifikasi peubah berpengaruh dengan SHAP	27
V SIMPULAN DAN SARAN	31
5.1 Simpulan	31
5.2 Saran	31
DAFTAR PUSTAKA	33
LAMPIRAN	35
RIWAYAT HIDUP	40



DAFTAR TABEL

1	<i>Confusion matrix</i> klasifikasi 4 kelas	9
2	Hubungan komponen <i>confusion matrix</i> untuk kelas ke- <i>k</i>	9
3	Peubah yang digunakan dalam penelitian	12
4	Kategorisasi kelas IP Semester 1 (skema 3 kelas)	13
5	Kategorisasi kelas IP Semester 1 (skema 4 kelas)	13
6	Nilai <i>hyperparameter</i> yang digunakan	13
7	Komposisi data pada setiap ulangan	18
8	Komposisi data latih sebelum dan sesudah penerapan <i>easyensemble</i>	19
9	<i>P-value</i> hasil ANOVA faktorial dua arah RF standar 3 kelas	21
10	<i>P-value</i> hasil ANOVA faktorial dua arah RF <i>easyensemble</i> 3 kelas	21
11	<i>P-value</i> hasil ANOVA faktorial dua arah RF standar 4 kelas	21
12	<i>P-value</i> hasil ANOVA faktorial dua arah RF <i>easyensemble</i> 4 kelas	22

DAFTAR GAMBAR

1	Diagram alir tahapan analisis data	14
2	Histogram sebaran IP Semester 1 mahasiswa D-IV jalur SNBP	15
3	<i>Boxplot</i> sebaran IP Semester 1 mahasiswa D-IV jalur SNBP berdasarkan program studi diterima	16
4	IP Semester 1 mahasiswa D-IV jalur SNBP (skema 3 kelas)	16
5	IP Semester 1 mahasiswa D-IV jalur SNBP (skema 4 kelas)	17
6	<i>Pair plot</i> korelasi <i>spearman</i>	17
7	<i>Boxplot Balanced Accuracy</i> dari (a) RF standar dan (b) RF <i>easyensemble</i> skema 3 kelas	23
8	<i>Boxplot Balanced Accuracy</i> dari (a) RF standar dan (b) RF <i>easyensemble</i> skema 4 kelas	23
9	<i>Boxplot Macro F1-Score</i> dari (a) RF standar dan (b) RF <i>easyensemble</i> skema 3 kelas	24
10	<i>Boxplot Macro F1-Score</i> dari (a) RF standar dan (b) RF <i>easyensemble</i> skema 4 kelas	24
11	<i>Boxplot G-Mean</i> dari (a) RF standar dan (b) RF <i>easyensemble</i> skema 3 kelas	25
12	<i>Boxplot G-Mean</i> dari (a) RF standar dan (b) RF <i>easyensemble</i> skema 4 kelas	25
13	Pohon keputusan tunggal RF <i>easyensemble</i> skema 3 kelas	27
14	SHAP <i>violin plot</i> Kelas 1	28
15	SHAP <i>violin plot</i> Kelas 2	28
16	SHAP <i>violin plot</i> Kelas 3	29



DAFTAR LAMPIRAN

1	Lampiran 1 Kriteria kategori buku perak	36
2	Lampiran 2 Daftar akronim program studi D-IV IPB University	36
3	Lampiran 3 Tabel ANOVA untuk Model 1	36
4	Lampiran 4 Tabel ANOVA untuk Model 2	37
5	Lampiran 5 Tabel ANOVA untuk Model 3	37
6	Lampiran 6 Tabel ANOVA untuk Model 4	37
7	Lampiran 7 <i>Confusion matrix</i> Model 1	37
8	Lampiran 8 Hasil uji asumsi ANOVA menggunakan uji <i>Shapiro-Wilk</i> dan uji <i>Levene</i>	38
9	Lampiran 9 Pohon keputusan tunggal RF <i>easyensemble</i> skema 3 kelas	39