

IDENTIFIKASI HUBUNGAN ANTARA MATA PELAJARAN SMA TERHADAP IP MAHASISWA DENGAN RANDOM FOREST

WARDATUS SYIFA



PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS, 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 Hubungan Antara Mata Pelajaran SMA Terhadap IP Mahasiswa dengan Random Forest” 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 Google Gemini 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

Wardatus Syifa
G1401221098



ABSTRAK

WARDATUS SYIFA. Identifikasi Hubungan Antara Mata Pelajaran SMA terhadap Indeks Prestasi Mahasiswa dengan Random Forest. Dibimbing oleh UTAMI DYAH SYAFITRI dan AAM ALAMUDI.

Transisi ke Kurikulum Merdeka memunculkan keragaman nilai mata pelajaran Sekolah Menengah Atas (SMA) yang menjadi tantangan dalam Seleksi Nasional Berdasarkan Prestasi (SNBP). Penelitian ini bertujuan untuk mengidentifikasi kontribusi nilai matematika SMA terhadap Indeks Prestasi (IP) semester satu mahasiswa sarjana jalur SNBP tahun 2025 serta menentukan *hyperparameter* model terbaik. Metode yang digunakan adalah klasifikasi *Random Forest* dengan *class weighting* untuk mengatasi ketidakseimbangan kelas, pengujian *hyperparameter* melalui *Analysis of Variance* (ANOVA) faktorial tiga arah, dan interpretasi peubah penting menggunakan *SHapley Additive exPlanations* (SHAP). Hasil penelitian menunjukkan bahwa model terbaik diperoleh pada skema pertama (3 kelas) dengan penerapan *class weight*, dengan nilai *F1-score* sebesar 42.99%, yang merupakan nilai tertinggi di antara seluruh skema yang diuji, serta didukung oleh *accuracy* sebesar 58.13%. Hal ini menunjukkan bahwa model tersebut memiliki keseimbangan terbaik antara *precision* dan *recall* dalam mengklasifikasikan IP mahasiswa. Selain itu, *max_depth* teridentifikasi sebagai *hyperparameter* yang paling berpengaruh signifikan terhadap performa model *Random Forest*. Berdasarkan analisis SHAP, rata-rata nilai matematika, nilai matematika semester lima, dan rata-rata nilai mata pelajaran pilihan teridentifikasi sebagai prediktor paling dominan dalam mengklasifikasikan IP mahasiswa. Temuan ini memberikan bukti empiris mengenai peran strategis penguasaan matematika jenjang menengah sebagai prediktor keberhasilan akademik awal di perguruan tinggi.

Kata kunci: ANOVA faktorial, indeks prestasi, matematika, Random Forest, SHAP



ABSTRACT

WARDATUS SYIFA. Identification of the Relationship Between High School Subjects and Student Grade Point Average Using Random Forest. Supervised by UTAMI DYAH SYAFITRI and AAM ALAMUDI.

The transition to the Merdeka Curriculum has created a diversity of high school subject grades, posing a challenge in the National Selection Based on Achievement (SNBP). This study aims to identify the contribution of high school mathematics grades to the first-semester Grade Point Average (GPA) of undergraduate students admitted through the 2025 SNBP and to determine the best model hyperparameters. The methods used were Random Forest classification with class weighting to handle class imbalance, hyperparameter testing via three-way factorial Analysis of Variance (ANOVA), and important variable interpretation using SHapley Additive exPlanations (SHAP). The results show that the best model was obtained using the first scheme (3 class) with class weight applied, achieving an F1-score of 42.99%, the highest among all tested schemes, supported by an accuracy of 58.13%. This indicates that the model achieved the best balance between precision and recall in classifying students' GPA. Furthermore, *max_depth* was identified as the hyperparameter with the most significant influence on the performance of the Random Forest model. Based on SHAP analysis, average mathematics scores, fifth-semester mathematics scores, and average elective course scores were identified as the most dominant predictors in classifying students' GPA. These findings provide empirical evidence regarding the strategic role of high-school-level mathematics proficiency as a predictor of early academic success in higher education.

Key words: Factorial ANOVA, Grade Point Average, Mathematics, Random Forest, SHAP



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

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.

Perpustakaan 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, 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.

IDENTIFIKASI HUBUNGAN ANTARA MATA PELAJARAN SMA TERHADAP IP MAHASISWA DENGAN RANDOM FOREST

WARDATUS SYIFA

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, 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. Agus Mohamad Soleh, S.Si., M.T



@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 Hubungan Antara Mata Pelajaran SMA terhadap IP
Mahasiswa dengan Random Forest

Nama : Wardatus Syifa

NIM : G1401221098

Disetujui oleh

Pembimbing 1:

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



Pembimbing 2:

Ir. Aam Alamudi, M.Si.



Diketahui oleh

Ketua Program Studi:

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

NIP. 19780411 200501 1002

Tanggal Ujian: 29 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 Juli 2026 ini ialah Indeks Prestasi (IP) dengan judul “Identifikasi Hubungan Mata Pelajaran SMA terhadap IP Mahasiswa dengan Random Forest”.

Terselesaikannya penulisan ini tidak lepas dari dukungan dan bantuan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada Ibu Dr. Utami Dyah Syafitri, S. Si., M.Si dan Bapak Ir. Aam Alamudi, M.Si, selaku dosen pembimbing. Terima kasih atas bimbingan, arahan, dan kesediaan meluangkan waktu di tengah kesibukan. Ucapan terima kasih juga ditujukan oleh Bapak Dr. Agus Mohamad Soleh, S.Si., M.T selaku dosen penguji ujian skripsi yang telah memberikan masukan dalam penulisan skripsi. Di samping itu, penghargaan penulis sampaikan kepada Direktorat Administrasi Pendidikan (DAPPMB) IPB yang telah memberikan serta mengizinkan penulis untuk menggunakan data mahasiswa dalam kepentingan penulisan karya ilmiah ini. Ungkapan terima kasih juga disampaikan kepada Bapak Wahono dan Ibu Yeti Mulyaningrum, orang tua yang selalu percaya dan mendo'akan saya.

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

Bogor, Agustus 2026

Wardatus Syifa



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	3
II TINJAUAN PUSTAKA	5
2.1 Indeks Prestasi (IP) Mahasiswa	5
2.2 <i>Random Forest</i>	5
2.3 <i>Class Weighting</i>	6
2.4 Rancangan Faktorial Tiga Arah	6
2.5 <i>SHapley Additive exPlanations</i> (SHAP)	8
2.6 Evaluasi Kebaikan Model	8
III METODE	11
3.1 Data	11
3.2 Prosedur Kerja	12
IV HASIL DAN PEMBAHASAN	15
4.1 Praproses Data	15
4.2 Eksplorasi Data	15
4.3 Partisi Data	21
4.4 Penanganan <i>Imbalanced Data</i> dan Pemodelan Random Forest	21
4.5 Pengujian <i>Hyperparameter</i> menggunakan Analysis of Variance (ANOVA) Faktorial Tiga Arah	24
4.6 Identifikasi Peubah Berpengaruh	27
V SIMPULAN DAN SARAN	32
5.1 Simpulan	32
5.2 Saran	32
DAFTAR PUSTAKA	33
LAMPIRAN	35
RIWAYAT HIDUP	40



DAFTAR TABEL

1	Tabel 1 <i>Confusion matrix</i> klasifikasi k kelas	9
2	Tabel 2 Rumus metrik evaluasi	10
3	Tabel 3 Daftar peubah dan penjelasan	11
4	Tabel 4 Klasifikasi IP kedua skema	12
5	Tabel 5 <i>Hyperparameter</i> yang digunakan	14
6	Tabel 6 Tabel Statistika deskriptif	16
7	Tabel 7 Partisi data skema pertama dan skema kedua	21
8	Tabel 8 Rata-rata metrik evaluasi data test	22
9	Tabel 9 <i>Confusion matrix fold</i> 1 skema pertama	22
10	Tabel 10 <i>Confusion matrix fold</i> 1 skema kedua	23
11	Tabel 11 F-hitung dan <i>p-value</i> hasil ANOVA	25
12	Tabel 12 Hasil uji asumsi normalitas	25
13	Tabel 13 Uji Asumsi Kehomogenan Ragam	26

DAFTAR GAMBAR

1	Gambar 1 Distribusi Data 3 Kelas	15
2	Gambar 2 Distribusi Data 4 Kelas	15
3	Gambar 3 Proporsi Klaster PPKU	17
4	Gmbar 4 Distribusi IP 3 Kelas	18
5	Gambar 5 Distribusi IP 4 Kelas	18
6	Gambar 6 Sebaran IP Berdasarkan Klaster PPKU	19
7	Gambar 7 Sebaran IP Berdasarkan Kode Program Studi	20
8	Gambar 8 Sebaran IP Mahasiswa SNBP	20
9	Gambar 9 SHAP	27
10	Gambar 10 SHAP Kelas 1	28
11	Gambar 11 SHAP Kelas 2	29
12	Gambar 12 SHAP Kelas 3	30
13	Gambar 13 <i>Heatmap</i> Korelasi Peubah Numerik	39

DAFTAR LAMPIRAN

1	Lampiran 1 Keterangan kode program studi	36
2	Lampiran 2 <i>Confusion matrix fold</i> 2 skema pertama	37
3	Lampiran 3 <i>Confusion matrix fold</i> 3 skema pertama	37
4	Lampiran 4 <i>Confusion matrix fold</i> 4 skema pertama	37
5	Lampiran 5 <i>Confusion matrix fold</i> 5 skema pertama	37
6	Lampiran 6 <i>Confusion matrix fold</i> 2 skema kedua	37
7	Lampiran 7 <i>Confusion matrix fold</i> 3 skema kedua	38
8	Lampiran 8 <i>Confusion matrix fold</i> 4 skema kedua	38
9	Lampiran 9 <i>Confusion matrix fold</i> 5 skema kedua	38
10	Lampiran 10 Korelasi spearmen peubah numerik	39