



# IDENTIFIKASI PEUBAH-PEUBAH SELEKSI YANG MEMENGARUHI PRESTASI MAHASISWA SARJANA ANGKATAN 2023 IPB UNIVERSITY MENGGUNAKAN *EXTRA-TREES*

**FEBRIAN ADHITYA CAHYA BELARDI**



**DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2024**



@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.



## PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Identifikasi Peubah-peubah Seleksi yang Memengaruhi Prestasi Mahasiswa Sarjana Angkatan 2023 IPB University Menggunakan *Extra-Trees*” 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, Agustus 2024

Febrian Adhitya Cahya Belardi  
G1401201041

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 mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@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.



## ABSTRAK

FEBRIAN ADHITYA CAHYA BELARDI. Identifikasi Peubah-peubah Seleksi yang Memengaruhi Prestasi Mahasiswa Sarjana Angkatan 2023 IPB University Menggunakan *Extra-Trees*. Dibimbing oleh ITASIA DINA SULVIANTI dan UTAMI DYAH SYAFITRI.

Transisi Seleksi Nasional Berbasis Prestasi (SNBP) oleh Kementerian Pendidikan dan Kebudayaan memicu penyesuaian internal di IPB University. Indeks Prestasi Kumulatif (IPK) merupakan metrik untuk mengukur prestasi mahasiswa. IPK dapat dipengaruhi oleh faktor multidimensional seperti aspek akademik dan latar belakang mahasiswa yang pada umumnya jenis datanya campuran antara kategorik dan numerik. Peubah-peubah seleksi SNBP yang jenis datanya campuran diduga memengaruhi IPK mahasiswa. Data IPK mahasiswa pada umumnya merupakan data tidak seimbang. *Supervised machine learning* seperti *Extra-Trees* memiliki kemampuan untuk mengklasifikasikan berbagai jenis data, salah satunya data seleksi SNBP tersebut. Penelitian ini bertujuan untuk mengidentifikasi peubah-peubah yang memengaruhi IPK mahasiswa di IPB University berdasarkan peubah-peubah seleksi SNBP 2023 menggunakan *Extra-Trees*. Data dibagi menjadi empat kelas IPK kemudian dipartisi menjadi data uji dan data latih dengan empat skema berbeda. Setiap data latih diseimbangkan dengan ADASYN dan tidak diseimbangkan, sehingga menghasilkan delapan set data. Model *Extra-Trees* dibangun dengan data latih menggunakan *GridSearchCV* untuk menentukan kombinasi *hyperparameter* terbaik. Model *Extra-Trees* pada data uji secara konsisten memiliki kinerja yang baik dalam memprediksi IPK mahasiswa dengan akurasi tertinggi mencapai 68,62% pada model tanpa proses ADASYN dengan skema partisi 70% data latih dan 30% data uji. Interpretasi *Shapley Additive exPlanation* (SHAP) pada model terbaik menunjukkan bahwa peubah paling berpengaruh terhadap IPK adalah program studi pilihan pertama mahasiswa saat mendaftar SNBP. Selain itu, prestasi, asal daerah, dan nilai mahasiswa semasa SMA juga termasuk peubah-peubah yang berpengaruh terhadap IPK.

Kata kunci: *extra-trees classifier*, indeks prestasi kumulatif, klasifikasi multikelas, *shaley additive explanation*

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.



## ABSTRACT

FEBRIAN ADHITYA CAHYA BELARDI. Identification of Selection Variabels Affecting the Academic Performance of Undergraduate Students of the 2023 Class at IPB University Using Extra-Trees. Supervised by ITASIA DINA SULVIANTI SUPERVISOR and UTAMI DYAH SYAFITRI.

Transition to the Achievement-Based National Selection by the Ministry of Education and Culture has prompted internal adjustments at IPB University. Grade Point Average (GPA), a key metric for assessing student achievement, can be influenced by multidimensional factors, including academic aspects and student background, which encompass both categorical and numerical data types. It is hypothesized that SNBP selection variables, characterized by mixed data types, impact student GPA. Notably, student GPA data tends to be unbalanced. Supervised machine learning algorithms, such as Extra-Trees, are proficient in classifying diverse data types, including SNBP selection data. This study aims to identify variables influencing student GPA at IPB University based on the 2023 SNBP selection variables using the Extra-Trees algorithm. The data were categorized into four GPA classes and partitioned into training and test sets using four distinct schemes. Each training dataset was balanced using ADASYN and left unbalanced, resulting in eight distinct datasets. The Extra-Trees models were constructed using GridSearchCV to optimize hyperparameter combinations. The Extra-Trees model consistently demonstrated robust performance in predicting student GPA on the test data, achieving the highest accuracy of 68.62% in the model without the ADASYN process under a partition scheme of 70% training data and 30% test data. Shapley Additive exPlanation (SHAP) analysis of the best model revealed that the most influential variable on GPA was the student's first-choice study program during SNBP registration. Additionally, academic achievements, regional origin, and high school grades were identified as significant variables influencing GPA.

**Keywords:** extra trees classifier, multiclass classification, shaley additive explanation, student grade point average



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 2024  
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-PEUBAH SELEKSI YANG MEMENGARUHI PRESTASI MAHASISWA SARJANA ANGKATAN 2023 IPB UNIVERSITY MENGGUNAKAN *EXTRA-TREES*

**FEBRIAN ADHITYA CAHYA BELARDI**

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

**DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2024**





**@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 Pika Silvianti, S.Si., 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 : Identifikasi Peubah-peubah Seleksi yang Memengaruhi Prestasi Mahasiswa Sarjana Angkatan 2023 IPB University Menggunakan *Extra-Trees*

Nama : Febrian Adhitya Cahya Belardi  
NIM : G1401201041

Disetujui oleh

Pembimbing 1:  
Dra. Itasia Dina Sulvianti, M.Si.



Pembimbing 2:  
Dr. Utami Dyah Syafitri, M.Si.



Diketahui oleh

Ketua Departemen Statistika:  
Dr. Bagus Sartono, S.Si., M.Si.  
NIP. 19780411 200501 1002

---

Tanggal Ujian:  
22 Juli 2024

Tanggal Lulus:

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 April 2024 sampai bulan Juli 2024 ini ialah Indeks Prestasi Kumulatif (IPK) Mahasiswa, dengan judul "Identifikasi Peubah-peubah Seleksi yang Memengaruhi Prestasi Mahasiswa Sarjana Angkatan 2023 IPB University Menggunakan *Extra-Trees*".

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

1. Bapak Murgunadi dan Ibu Pik Tiyaningsih selaku kedua orang tua serta keluarga yang tidak henti memberikan doa, dukungan dan motivasi;
2. Ibu Dra. Itasia Dina Sulvianti, M.Si. dan Ibu Dr. Utami Dyah Syafitri, M.Si. selaku dosen pembimbing yang telah bersedia untuk membimbing dan memberikan saran serta masukan kepada penulis sehingga penyusunan karya ilmiah ini dapat diselesaikan dengan baik;
3. Ibu Pika Silvianti, S.Si., M.Si. selaku dosen penguji, Bapak Aam Alamudi, M.Si. selaku moderator seminar hasil, dan Bapak Septian Rahardiantoro, S.Stat., M.Si., Ph.D selaku moderator kolokium yang telah memberikan saran dan masukan dalam penulisan karya ilmiah ini;
4. dosen dan tenaga pendidik Departemen Statistika IPB University yang telah memberikan ilmu yang bermanfaat dalam membantu kebutuhan administrasi penulis;
5. Direktorat Administrasi Pendidikan IPB University yang telah memberikan penulis kesempatan untuk menggunakan data mahasiswa dalam penelitian ini;
6. sahabat-sahabat sepermainan penulis yang telah meluangkan waktunya untuk memberikan motivasi dan dukungan ketika penulis merasa jenuh serta lelah;
7. teman-teman Statistika angkatan 57 yang telah kebersamai penulis dalam menjalani kehidupan perkuliahan; dan
8. seluruh pihak terkait yang telah membantu penulis dalam menyusun karya ilmiah ini yang tidak dapat penulis sebutkan satu persatu

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan. Penulis menyadari masih terdapat kesalahan dan kekurangan penyusunan karya ilmiah ini. Semoga Allah subhanaahu wa ta'ala selalu memberikan rahmat dan hidayah-Nya kepada kita semua.

Bogor, Agustus 2024

*Febrian Adhitya Cahya Belardi*

## DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	x
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 IPK (Indeks Prestasi Kumulatif)	3
2.2 <i>Extremely Randomized Trees (Extra-Trees)</i>	3
2.3 <i>Adaptive Syntethic Sampling (ADASYN)</i>	5
2.4 <i>Shapley Additive Explanations (SHAP)</i>	6
2.5 Evaluasi Kebaikan Model	7
III METODE	9
3.1 Data	9
3.2 Prosedur Analisis Data	10
IV HASIL DAN PEMBAHASAN	14
4.1 Praproses Data	14
4.2 Eksplorasi Data dan Rekayasa Peubah	14
4.3 Partisi Data dan Penanganan Data Tidak Seimbang	16
4.4 Klasifikasi dan Evaluasi	17
4.5 Ukuran Kepentingan Peubah	21
V SIMPULAN DAN SARAN	23
5.1 Simpulan	23
5.2 Saran	23
DAFTAR PUSTAKA	24
LAMPIRAN	27
RIWAYAT HIDUP	31



## DAFTAR TABEL

1	<i>Confusion matrix</i> pada klasifikasi 4 kelas	7
2	Peubah respon dan penjelas yang digunakan dalam penelitian	9
3	<i>Hyperparameter</i> yang digunakan	12
4	<i>Confusion matrix</i> peubah hasil rekayasa	15
5	Peubah baru hasil rekayasa	16
6	Skema <i>hyperparameter tuning</i>	18
7	Evaluasi hasil model klasifikasi dengan <i>Extra-Trees</i>	19
8	<i>Confusion matrix</i> hasil prediksi dari model 4	21

## DAFTAR GAMBAR

1	Ilustrasi algoritma <i>Extra-Trees</i>	4
2	Diagram alir analisis data pada penelitian	13
3	Histogram (a) dan <i>Bar Chart</i> (b) data IPK mahasiswa IPB University angkatan 2023	14
4	Proporsi banyaknya amatan (a) sebelum ADASYN dan (b) sesudah ADASYN	17
5	Sebaran akurasi setiap kombinasi <i>hyperparameter</i> di masing-masing	19
6	<i>Summary plot</i> (a) dan <i>barplot</i> (b) tingkat kepentingan peubah pada IPK	22

## DAFTAR LAMPIRAN

1	<i>Confusion matrix</i> hasil prediksi dari model 1	28
2	<i>Confusion matrix</i> hasil prediksi dari model 2	28
3	<i>Confusion matrix</i> hasil prediksi dari model 3	28
4	<i>Confusion matrix</i> hasil prediksi dari model 5	28
5	<i>Confusion matrix</i> hasil prediksi dari model 6	28
6	<i>Confusion matrix</i> hasil prediksi dari model 7	29
7	<i>Confusion matrix</i> hasil prediksi dari model 8	29
8	Nilai <i>mean( SHAP Value )</i> untuk semua peubah	29