

PENANGANAN DATA TIDAK SEIMBANG MODEL *RANDOM FOREST* DAN *XGBOOST* BERBASIS OPTIMASI *BAYESIAN* PADA KLASIFIKASI IPK PROGRAM MAGISTER

ARIFUDDIN R



**PROGRAM STUDI STATISTIKA DAN SAINS DATA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**

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 TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Penanganan Data Tidak Seimbang Model *Random Forest* dan *XGBoost* Berbasis Optimasi *Bayesian* pada Klasifikasi IPK Program Magister” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apapun 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 tesis ini.

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

Bogor, Juli 2024

Arifuddin R
G1501201019

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.



RINGKASAN

ARIFUDDIN R. Penanganan Data Tidak Seimbang Model *Random Forest* dan *XGBoost* Berbasis Optimasi *Bayesian* pada Klasifikasi IPK Program Magister. Dibimbing oleh UTAMI DYAH SYAFITRI dan ERFIANI.

Dalam upaya menghasilkan Sumber Daya Manusia (SDM) unggul, Program Magister SPs-IPB perlu melakukan proses seleksi yang ketat pada penerimaan mahasiswa baru. Indeks Prestasi Kumulatif (IPK) sebagai indikator untuk melihat prestasi akademik mahasiswa, umumnya berkorelasi positif terhadap kualitas SDM unggul. Salah satu metode yang dapat digunakan dalam mengelompokkan IPK adalah model klasifikasi *machine learning*, hasil klasifikasi bisa menjadi acuan pada proses seleksi mahasiswa. Metode klasifikasi *machine learning* saat ini telah dikembangkan pada berbagai bidang, salah satunya bidang pendidikan. Algoritma *Random Forest* dan *XGBoost* merupakan model *ensemble learning* yang sering digunakan pada kasus klasifikasi, karena umumnya menghasilkan akurasi yang tinggi. Algoritma ini mampu beradaptasi dengan data yang besar dan kompleks seperti pada model *nonlinear*.

Salah satu tahapan penting dalam membangun model *machine learning* adalah mencari *hyperparameter* yang dapat menghasilkan model optimal, umumnya dikenal dengan istilah optimasi atau *tuning hyperparameter*. Hal ini dikarenakan *hyperparameter* secara langsung mengontrol proses pelatihan algoritma, sehingga memiliki efek signifikan pada performa model. Optimasi *Bayesian* dapat menjadi alternatif yang lebih efektif dalam pencarian *hyperparameter* optimal. Metode ini menggunakan contoh *hyperparameter* yang nantinya akan dioptimalisasi, hingga menghasilkan model dengan nilai metrik paling baik. Optimasi *Bayesian* mampu memberikan hasil yang lebih baik dan efisien dibandingkan metode seperti *babysitting*, *grid search* dan *random search*.

Berbagai macam permasalahan data yang dapat ditemui pada kasus klasifikasi, salah satunya adalah ketidakseimbangan data. Kondisi ini terjadi ketika data memiliki sebaran amatan antar kelas yang berbeda jauh. Pada kasus data tidak seimbang, umumnya kelas mayoritas menghasilkan prediksi yang bias. Model akan fokus mempelajari kategori dengan amatan terbanyak, sehingga kelas minoritas sering kali diprediksi sebagai kelas mayoritas. Penanganan data tidak seimbang yang terbukti mampu meningkatkan prediksi kelas minoritas adalah membangkitkan amatan sintetis. Metode yang dapat digunakan seperti SMOTE, SMOTE for Nominal Continuous (SMOTE-NC) dan ADASYN, tiga metode ini menghasilkan data sintetis yang lebih bervariasi dibandingkan *random oversampling*.

Berdasarkan permasalahan yang telah dipaparkan, penelitian ini bertujuan untuk mengklasifikasikan IPK mahasiswa program magister berdasarkan kategori unggul ($IPK \geq 3,5$) dan non unggul ($IPK < 3,5$). Algoritma yang digunakan adalah *Random Forest* dan *XGBoost* dengan optimasi *Bayesian* dan *random search*. Pada penanganan data tidak seimbang, digunakan tiga perbandingan metode yaitu SMOTE, SMOTE-NC dan ADASYN. Data yang digunakan terdiri atas empat angkatan (2020-2023) dengan jumlah amatan 4677 mahasiswa. Model dibangun



berdasarkan data tahun berbeda, serta diaplikasikan pada data tahun selanjutnya untuk melihat relevansi hasil ketika diaplikasikan pada data baru.

Hasil yang diperoleh menunjukkan tiga metode penanganan data tidak seimbang mampu meningkatkan nilai *sensitivity* model. Model dengan penanganan data tidak seimbang memiliki performa yang lebih baik dalam mengklasifikasikan kelas minoritas (IPK *non* unggul). Model terbaik diperoleh dari penanganan SMOTE-NC menggunakan optimasi *Bayesian* yang terbukti lebih efisien secara waktu dibandingkan *random search*. Peubah yang paling berpengaruh pada perolehan IPK mahasiswa adalah X8 (Status PT Asal), X2 (IPK S1), X6 (Asal Perguruan Tinggi), X1 (Usia) dan X11 (Program Studi). Peubah X4 (Gender), X9 (Jalur Masuk) dan X5 (Status Pernikahan) cenderung tidak berpengaruh berdasarkan model terbaik. Metrik hasil klasifikasi data tahun selanjutnya relatif sama, sehingga peubah maupun model klasifikasi yang digunakan masih relevan diaplikasikan pada proses seleksi mahasiswa.

Kata kunci: Indeks Prestasi Kumulatif (IPK), optimasi Bayesian, penanganan data tidak seimbang, Random Forest, 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.

SUMMARY

ARIFUDDIN R. Handling Imbalanced Data in Random Forest and XGBoost Models with Bayesian Optimization-Based Method for Master's Program GPA Classification. Supervised by UTAMI DYAH SYAFITRI and ERFIANI.

To produce excellent human resources, the IPB University Graduate School Master's Program must carry out a strict selection process for accepting new students. The Grade Point Average (GPA), as an indicator of student academic achievement, generally correlates positively with the quality of excellent human resources. One method that can be used to classify GPA is a machine learning classification model, the classification results can be used as a reference in the student selection process. Machine learning classification methods have currently been developed in various fields, one of which is education. The Random Forest and XGBoost algorithms are ensemble learning models often used in classification cases because they generally produce high accuracy. This algorithm can adapt to large and complex data, such as nonlinear models.

One of the critical stages in developing a machine learning model is the search for hyperparameters that can produce an optimal model, a process known as optimization or hyperparameter tuning. The choice of hyperparameters significantly influences the model's performance. In this context, Bayesian optimization emerges as a more efficient alternative for finding optimal hyperparameters. This method utilizes example hyperparameters, which are then optimized to produce a model with the best metric values. Compared to traditional methods like babysitting, grid search, and random search, Bayesian optimization consistently delivers superior and more efficient results.

Data problems can be encountered in classification cases, such as data imbalance. This condition occurs when the data has a very different distribution of observations between classes. In the case of imbalanced data, the majority class generally produces biased predictions. The model will study the categories with the most observations so that the minority class is often predicted as the majority class. Handling imbalanced data that is proven to be able to improve minority class predictions is generating synthetic observations. Methods that can be used include SMOTE, SMOTE for Nominal Continuous (SMOTE-NC) and ADASYN; these methods produce synthetic data that is more varied than random oversampling.

Based on the problems that have been explained, this research will compare three methods for handling unbalanced data, namely SMOTE, SMOTE-NC and ADASYN, to classify the GPA of Master's Program students based on excellent ($GPA \geq 3.5$) and non-excellent ($GPA < 3.5$) categories. The algorithms used are Random Forest and XGBoost with Bayesian and random search optimization. The data comprises four classes (2020-2023) with 4677 students observed. The model is built based on data from different years and applied to the following year's data to see the relevance of the results when applied to new data.

The results show that three methods for handling unbalanced data can increase the sensitivity value of the model so that it performs better in classifying minority classes (Non-excellent GPA). The best model was obtained from handling SMOTE-NC using Bayesian optimization, which was more time efficient than random search. The variables that have the most influence on students' GPA are X8



(Undergraduate University Status), X2 (Undergraduate GPA), X6 (Undergraduate University), X1 (Age), and X11 (Study Program). In contrast, variables X4 (Gender), X9 (Enrolment), and X5 (Marital Status) tend to have no effect. The metrics for the following year's data classification results are relatively the same, so the variables and classification models used are still relevant to applying the student selection process.

Keywords: Bayesian optimization, Grade Point Average (GPA), Handling imbalanced data, Random Forest, 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 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 2024¹
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.



@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.

PENANGANAN DATA TIDAK SEIMBANG MODEL *RANDOM FOREST* DAN *XGBOOST* BERBASIS OPTIMASI BAYESIAN PADA KLASIFIKASI IPK PROGRAM MAGISTER

ARIFUDDIN R

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI STATISTIKA DAN SAINS DATA
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.

Penguji Luar Komisi Pembimbing pada Ujian Tesis:
Dr. Agus M. Soleh, S.Si., M.T.



Judul Tesis : Penanganan Data Tidak Seimbang Model *Random Forest* dan *XGBoost* Berbasis Optimasi *Bayesian* pada Klasifikasi IPK
Program Magister
Nama : Arifuddin R
NIM : G1501201019

Disetujui oleh

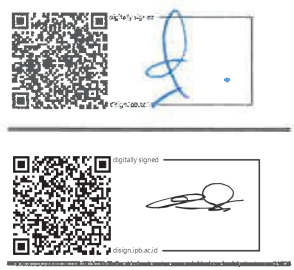
Pembimbing 1:
Dr. Utami Dyah Syafitri, S.Si., M.Si

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



Diketahui oleh

Ketua Program Studi:
Dr. Agus M. Soleh, S.Si., M.T.
NIP 197503151999031004
Dekan Fakultas Matematika dan
Ilmu Pengetahuan Alam:
Dr. Berry Juliandi, S.Si., M.Si.
NIP 197807232007011001



Tanggal Ujian:
27 Juni 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 ini ialah klasifikasi menggunakan model *machine learning* dengan judul “Penanganan Data Tidak Seimbang Model *Random Forest* dan *XGBoost* Berbasis Optimasi *Bayesian* pada Klasifikasi IPK Program Magister”.

Semoga karya ilmiah ini dapat memberikan manfaat bagi pihak yang membutuhkan dan berkontribusi pada kemajuan ilmu pengetahuan. Penulis menyadari bahwa penyusunan penelitian ini terselesaikan dengan baik karena adanya bantuan, dorongan, dan bimbingannya dari banyak pihak. Oleh karena itu, ucapan terima kasih penulis sampaikan kepada:

1. Ibu Dr. Utami Dyah Syafitri, S.Si., M.Si. dan Ibu Dr. Ir. Erfiani, M.Si. selaku komisi pembimbing yang senantiasa bersedia meluangkan waktunya untuk membimbing, memberikan saran, serta memberikan masukan kepada penulis sehingga penelitian ini dapat diselesaikan dengan baik.
2. Bapak Dr. Agus M. Soleh, S.Si., M.T. selaku penguji luar komisi yang telah banyak memberikan masukan dan saran dalam penulisan tesis ini.
3. Pimpinan, pengajar, dan pegawai Program Studi Statistika dan Sains Data yang telah banyak membantu pada proses penyelesaian studi dan penelitian ini.
4. Lembaga Pengelola Dana Pendidikan (LPDP) yang telah menjadi sponsor penulis dalam melanjutkan pendidikan pada jenjang magister di Program Studi Statistika dan Sains Data IPB University.
5. Kak Dr. Anwar Lubis, M.Si dari Universitas Negeri Makassar yang telah banyak membantu penulis dalam proses penyelesaian studi di IPB University.
6. Sahabat, teman, dan rekan-rekan mahasiswa pascasarjana Program Studi Statistika dan Sains Data atas bantuan dan kebersamaannya.
7. Teristimewa untuk Ayah (Alm), Ibu, kakak, dan seluruh keluarga yang selalu mendukung, mendoakan, dan memotivasi penulis dalam proses dan penyelesaian studi ini.

Bogor, Juli 2024

Arifuddin R

DAFTAR ISI

DAFTAR TABEL	viii
DAFTAR GAMBAR	viii
DAFTAR LAMPIRAN	ix
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
II TINJAUAN PUSTAKA	4
2.1 Penanganan Data Tidak Seimbang	4
2.2 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	5
2.3 <i>Synthetic Minority Oversampling Technique for Nominal Continuous (SMOTE-NC)</i>	5
2.4 <i>Adaptive Synthetic Resampling (ADASYN)</i>	6
2.5 Optimasi <i>Bayesian</i>	7
2.6 <i>Random Forest (RF)</i>	8
2.7 <i>Extreme Gradient Boosting (XGBoost)</i>	9
2.8 Metrik Evaluasi Performa Model	10
III METODE	11
3.1 Data Penelitian	11
3.2 Tahapan Analisis Data	12
IV HASIL DAN PEMBAHASAN	15
4.1 Persiapan dan Eksplorasi Data	15
4.2 Penanganan Data Tidak Seimbang	21
4.3 Optimasi <i>Hyperparameter</i>	22
4.4 Evaluasi Hasil Klasifikasi <i>Random Forest</i> dan <i>XGBoost</i>	25
4.5 Tingkat Kepentingan Peubah	27
V SIMPULAN DAN SARAN	30
5.1 Simpulan	30
5.2 Saran	30
DAFTAR PUSTAKA	31
LAMPIRAN	34
RIWAYAT HIDUP	43



DAFTAR TABEL

1	Data yang digunakan pada contoh aplikasi SMOTE-NC	6
2	Contoh tabel <i>confusion matrix</i> dua kelas	10
3	Daftar peubah prediktor yang digunakan membangun model klasifikasi	11
4	Kombinasi model klasifikasi <i>Random Forest</i> dan <i>XGBoost</i> berdasarkan metode optimasi dan penanganan data tidak seimbang	14
5	Jumlah data hasil <i>splitting</i> dan data sintetis setelah penanganan data tidak seimbang	22
6	Domain <i>hyperparameter</i> model <i>Random Forest</i> dan <i>XGBoost</i>	22
7	Perbandingan metrik kinerja model <i>Random Forest</i> berdasarkan metode penanganan data tidak seimbang	26
8	Perbandingan metrik kinerja model <i>XGBoost</i> berdasarkan metode penanganan data tidak seimbang	26
9	Perbandingan Kombinasi nilai <i>hyperparameter</i> model terbaik dari <i>Random Forest</i> dan <i>XGBoost</i> pada setiap penanganan data tidak seimbang	27
10	Hasil klasifikasi data tahun $i+1$ menggunakan model terbaik tahun i	28

DAFTAR GAMBAR

1	Ilustrasi penambahan amatan sintetis pada penanganan data tidak seimbang menggunakan jarak tetangga terdekat kelas minoritas	4
2	Ilustrasi pada proses optimasi <i>Bayesian</i>	8
3	Ilustrasi <i>Random Forest</i>	9
4	Diagram alir tahapan analisis data	12
5	Persentase kategori perolehan IPK: (a) tahun 2020, (b) tahun 2021, dan (c) tahun 2022	15
6	Gambaran profil mahasiswa berdasarkan peubah kategorik data tahun 2020: (a) <i>barchart</i> persentase mahasiswa setiap kategori dan (b) <i>stacked bar</i> persentase perolehan IPK setiap kategori	15
7	Gambaran mahasiswa berdasarkan peubah kategorik data tahun 2021: (a) <i>barchart</i> persentase mahasiswa setiap kategori dan (b) <i>stacked bar</i> persentase perolehan IPK setiap kategori	16
8	Gambaran mahasiswa berdasarkan peubah kategorik data tahun 2022: (a) <i>barchart</i> persentase mahasiswa setiap kategori dan (b) <i>stacked bar</i> persentase perolehan IPK setiap kategori	17
9	Gambaran mahasiswa berdasarkan peubah kategorik data tahun 2023: (a) <i>barchart</i> persentase mahasiswa setiap kategori dan (b) <i>stacked bar</i> persentase perolehan IPK setiap kategori	18
10	Nilai koefisien <i>silhouette</i> pada penentuan jumlah kluster optimal	19
11	Dendogram hasil penggerombolan program studi menggunakan metode pautan lengkap (<i>complete linkage</i>): jumlah kluster dua (a) dan jumlah kluster tiga (b)	20

12	Diagram kotak garis karakteristik klaster program studi yang terbentuk berdasarkan peubah Usia, IPK S1 dan <i>Gap Year</i>	21
13	Ilustrasi split data <i>training</i> pada proses validasi silang	23
14	Diagram kotak garis sebaran nilai metrik evaluasi model <i>Random Forest</i> (2020-2022) pada data <i>training</i> dengan penanganan data tidak seimbang menggunakan dua metode optimasi	24
15	Diagram kotak garis sebaran nilai metrik evaluasi model <i>XGBoost</i> (2020-2022) pada data <i>training</i> dengan penanganan data tidak seimbang menggunakan dua metode optimasi	25
16	Peubah yang berpengaruh pada perolehan IPK berdasarkan model: (a) <i>Random Forest</i> dan (b) <i>XGBoost</i>	29

DAFTAR LAMPIRAN

1	Lampiran 1 Data jumlah mahasiswa Program Magister SPs-IPB dari tahun 2020 sampai 2023	35
2	Lampiran 2 Hasil uji <i>chi square</i> peubah prediktor tipe kategorik terhadap peubah respon data tahun 2020, 2021, dan 2022	36
3	Lampiran 3 Anggota dan nilai pusat setiap klaster yang terbentuk	37
4	Lampiran 4 Ilustrasi penambahan amatan sintetis pada kelas minoritas menggunakan tiga metode penanganan data tidak seimbang	37
5	Lampiran 5 Eksplorasi awal jumlah pohon yang akan dibangun pada model <i>Random Forest</i> data tahun 2020, 2021, dan 2022	38
6	Lampiran 6 Perbandingan waktu (menit) dalam pencarian model terbaik <i>Random Forest</i> dan <i>XGBoost</i> menggunakan optimasi <i>Bayesian</i> dan <i>random search</i> (RS)	40
7	Lampiran 7 Eksplorasi awal jumlah pohon yang akan dibangun pada model <i>XGBoost</i> data tahun 2020, 2021, dan 2022	41