



PENERAPAN *RANDOM FOREST* DALAM MENKLASIFIKASIKAN KETEPATAN WAKTU KELULUSAN MAHASISWA PROGRAM SARJANA IPB UNIVERSITY SEMASA PANDEMI

ARIE WIJAYA



**DEPARTEMEN STATISTIKA
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

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 “Penerapan *Random Forest* Dalam Mengklasifikasikan Ketepatan Waktu Kelulusan Mahasiswa Program Sarjana IPB University Semasa Pandemi” 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

Arie Wijaya
G14190081



ABSTRAK

ARIE WIJAYA. Penerapan *Random Forest* Dalam Mengklasifikasikan Ketepatan Waktu Kelulusan Mahasiswa Program Sarjana IPB University Semasa Pandemi. Dibimbing oleh RAHMA ANISA dan GERRY ALFA DITO.

Lulus tepat waktu merupakan salah satu indikator kualitas lulusan dalam menyelesaikan studinya. Mahasiswa yang lulus tepat waktu adalah mahasiswa yang dapat menyelesaikan studinya dalam kurun waktu ≤ 8 semester. Penelitian sebelumnya menunjukkan bahwa berdasarkan data kelulusan mahasiswa tahun 2010 hingga 2017, persentase mahasiswa yang lulus tepat waktu masih sangat rendah berkisar 30%. Upaya untuk meningkatkan rasio lulus tepat waktu dapat dilakukan dengan analisis pemodelan menggunakan *random forest* untuk mengidentifikasi faktor-faktor yang mempengaruhi. *Random forest* dipilih karena mampu memberikan performa yang baik, mudah diimplementasikan, dan efektif dalam menangani data yang kompleks. Namun data yang digunakan dalam penelitian ini cenderung memiliki masalah ketidakseimbangan data yang ditangani menggunakan *Synthetic Minority Oversampling Technique* (SMOTE). Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja model dari hasil analisis *random forest* dengan SMOTE dan tanpa SMOTE dan mengidentifikasi peubah penting dalam mengklasifikasikan ketepatan waktu kelulusan mahasiswa program sarjana IPB University. Hasil penelitian menunjukkan bahwa performa model dengan penanganan SMOTE lebih baik dibandingkan tanpa penanganan SMOTE. Peubah penting dalam penelitian adalah IPK terakhir dan IPK PKU, dimana peubah tersebut sangat mempengaruhi ketepatan waktu kelulusan mahasiswa.

Kata kunci : data tidak seimbang, klasifikasi, lulus tepat waktu, *random forest*, SMOTE.

ABSTRACT

ARIE WIJAYA. Application of Random Forest in Classifying the Timeliness of Graduation for Undergraduate Students at IPB University During the Pandemic. Supervised by RAHMA ANISA and GERRY ALFA DITO.

Graduating on time is one of the quality indicators for graduates in completing their studies. Students who graduate on time are students who can complete their studies within ≤ 8 semesters. Previous research shows that based on student graduation data from 2010 to 2017, the percentage of students who graduated on time was around 30%. Efforts to increase the on-time graduation ratio can be done by modeling analysis using random forest to identify influencing factors. Random forest was chosen because it can provide good performance, easy to implement, and effective in handling complex data. However, the data used in this study tend to have an imbalance data problem was handled using the Synthetic Minority Oversampling Technique (SMOTE). Therefore, this research aims to compare the model performance from the results of random forest analysis with SMOTE and without SMOTE and to identify variables importance to classify the timeliness of graduation of IPB University undergraduate students. The results of this research showed that the performance of the model with SMOTE handling is better than without SMOTE. The important variables in the research were the final GPA and PKU GPA, where these variables greatly influenced the timeliness of students graduation.

Keywords : classification, graduating on time, imbalance data, random forest, SMOTE.



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.



PENERAPAN *RANDOM FOREST* DALAM MENGLASIFIKASIKAN KETEPATAN WAKTU KELULUSAN MAHASISWA PROGRAM SARJANA IPB UNIVERSITY SEMASA PANDEMI

ARIE WIJAYA

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:
Ir. Aam Alamudi, M.Si.**



@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.



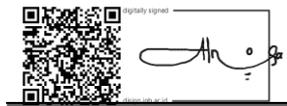
Judul Skripsi : Penerapan *Random Forest* Dalam Mengklasifikasikan Ketepatan Waktu Kelulusan Mahasiswa Program Sarjana IPB University Semasa Pandemi

Nama : Arie Wijaya
NIM : G14190081

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:
Rahma Anisa, S.Stat, M.Si.



Pembimbing 2:
Gerry Alfa Dito, S.Si, M.Si.



Diketahui oleh

Ketua Departemen Statistika:
Dr. Bagus Sartono, S.Si, M.Si.
NIP 197804112005011002



Tanggal Ujian:
(2 Agustus 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 Februari 2023 sampai bulan Juli 2024 ini ialah klasifikasi, dengan judul “Penerapan *Random Forest* Dalam Mengklasifikasikan Ketepatan Waktu Kelulusan Mahasiswa Program Sarjana IPB University Semasa Pandemi”. Terima kasih penulis ucapkan kepada :

1. Ibu Rahma Anisa, S.Stat, M.Si, dan Bapak Gerry Alfa Dito, S.Si, M.Si, selaku dosen pembimbing yang telah senantiasa membimbing, memberi saran dan masukan dalam penulisan karya ilmiah ini.
2. Bapak Ir. Aam Alamudi, M.Si. selaku dosen penguji yang telah memberikan kritik dan saran dalam penulisan karya ilmiah ini.
3. Seluruh dosen pengajar Departemen Statistika IPB yang telah memberikan penulis banyak ilmu dan pembelajaran bermanfaat selama perkuliahan.
4. Bapak M. Ali Irfani dan Ibu Siti Hajir selaku orang tua penulis, Arief Effandi, Rissa Haryanti, Lies Kurniawati, Andika Setiawan, dan Novira Hidayati selaku kakak dan abang penulis yang senantiasa memberikan doa, dukungan, motivasi, nasihat, dan kasih sayang bagi penulis.
5. Seluruh Staff Tata Usaha Departemen Statistika IPB yang telah banyak membantu penulis dalam proses administrasi perkuliahan.
6. Anggun Andini, Elisabeth Kurniasari, Grashella Clara, Resti Hadayani, Rochmat Fahrul Rozzy, serta teman-teman statistika angkatan 56 lainnya dan teman-teman HIMAJA dan KKN-T yang telah memberikan banyak bantuan, saran, motivasi, serta semangat selama masa perkuliahan hingga penyusunan karya ilmiah ini.
7. Seluruh pihak terkait yang tidak dapat penulis sebutkan satu persatu atas segala doa, kebaikan, dukungan, serta motivasi sehingga penulis dapat menyelesaikan karya ilmiah ini.

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

Bogor, Agustus 2024

Arie Wijaya



DAFTAR ISI

DAFTAR TABEL	
DAFTAR GAMBAR	
DAFTAR LAMPIRAN	
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 <i>Random Forest</i>	3
2.2 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	4
2.3 <i>K-Fold Cross Validation</i>	6
2.4 Ukuran Keباikan Model	6
2.5 <i>Mean Decrease Gini (MDG)</i>	7
III METODE	9
3.1 Data	9
3.2 Prosedur Analisis	10
IV HASIL DAN PEMBAHASAN	12
4.1 Pra-Proses Data	12
4.2 Eksplorasi Data	13
4.3 Penanganan <i>Imbalance Data</i> (Data Tidak Seimbang)	19
4.4 Analisis <i>Random Forest</i>	19
4.5 Evaluasi Model	20
4.6 Peubah Penting (<i>Variable Importance</i>)	21
V SIMPULAN DAN SARAN	23
5.1 Simpulan	23
5.2 Saran	23
DAFTAR PUSTAKA	24
LAMPIRAN	27
RIWAYAT HIDUP	42



DAFTAR TABEL

1	Matriks Konfusi	7
2	Data peubah respon dan peubah penjelas	9
3	Data peubah respon dan peubah penjelas setelah dilakukan pelabelan	10
4	Distribusi data hilang berdasarkan peubah	12
5	Nilai <i>hyperparameter</i> yang digunakan	19
6	Perbandingan hasil model <i>random forest</i> dengan SMOTE dan tanpa SMOTE	20

DAFTAR GAMBAR

1	Ilustrasi algoritma <i>random forest</i>	3
2	Persentase mahasiswa berdasarkan ketepatan waktu kelulusan	13
3	Persentase mahasiswa berdasarkan fakultas dan ketepatan waktu kelulusan	13
4	Persentase mahasiswa berdasarkan jenis kelamin dan ketepatan waktu kelulusan	14
5	Persentase mahasiswa berdasarkan jalur masuk dan ketepatan waktu kelulusan	15
6	Persentase mahasiswa berdasarkan asal sekolah dan ketepatan waktu kelulusan	15
7	Persentase mahasiswa berdasarkan kategori sekolah dan ketepatan waktu kelulusan	16
8	Persentase mahasiswa berdasarkan pengambilan minor dan ketepatan waktu kelulusan	16
9	Persentase mahasiswa berdasarkan penerimaan beasiswa dan ketepatan waktu kelulusan	17
10	<i>Boxplot</i> hubungan antara IPK PKU dan ketepatan waktu lulus	18
11	<i>Boxplot</i> hubungan antara IPK terakhir dan ketepatan waktu lulus	18
12	Nilai <i>mean decrease gini</i> peubah penjelas	21

DAFTAR LAMPIRAN

1	Rata-rata IPK pada setiap peubah	28
2	Nilai akurasi tanpa SMOTE pada <i>hyperparameter tuning</i>	30
3	Nilai akurasi dengan SMOTE pada <i>hyperparameter tuning</i>	35

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.