



PEMODELAN KLASIFIKASI KETERLAMBATAN PEMBAYARAN UKT MAHASISWA IPB DENGAN RANDOM FOREST DAN ADABOOST

FAREL FIRMAN



**DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2021**



@Hak cipta milik IPBUniversity

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 IPBUniversity.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPBUniversity.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi berjudul *Pemodelan Klasifikasi Keterlambatan Pembayaran UKT Mahasiswa IPB dengan Random Forest dan AdaBoost* adalah benar karya saya dengan arahan dari komisi 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 skripsi ini.

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

Bogor, Februari 2021

Farel Firman
NIM G14160046

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 IPBUniversity.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPBUniversity.



@Hak cipta milik IPBUniversity

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 IPBUniversity.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPBUniversity.



ABSTRAK

FAREL FIRMAN. Pemodelan Klasifikasi Keterlambatan Pembayaran UKT Mahasiswa IPB dengan *Random Forest* dan *AdaBoost*. Dibimbing oleh MOHAMMAD MASJKUR dan CICI SUHAENI.

Menurut data yang diperoleh dari data Direktorat Keuangan dan Akutansi IPB, terjadi peningkatan jumlah mahasiswa yang mengalami keterlambatan dalam pembayaran UKT di tiap angkatan mulai tahun 2016 sampai dengan 2018 dengan banyak mahasiswa yang mengalami keterlambatan yaitu 54, 64, dan 117 mahasiswa. Penelitian ini bertujuan untuk membandingkan pemodelan klasifikasi menggunakan *Random Forest* dan *AdaBoost* serta mengidentifikasi peubah penting yang didapatkan dari pemodelan terbaik. Pemodelan klasifikasi akhir *Random Forest* menggunakan parameter optimal banyak peubah penjelas sebanyak 7 peubah dan banyak pohon sebanyak 500 pohon dengan penanganan ketidakseimbangan data yaitu *Random Oversampling* (ROS). Sedangkan pemodelan klasifikasi akhir *AdaBoost* menggunakan parameter optimal banyak iterasi sebanyak 80 iterasi dengan penanganan ketidakseimbangan data yaitu *Random Undersampling* (RUS). Pemodelan *Random Forest* dan *AdaBoost* memiliki ukuran kinerja prediksi ROC-AUC masing-masing sebesar 58,70% dan 52,90%. Pemodelan *Random Forest* memiliki kinerja prediksi lebih baik dibandingkan dengan *AdaBoost* dalam memprediksi mahasiswa yang berindikasi akan terlambat dalam pembayaran UKT dengan selisih nilai ROC-AUC sebesar 5,8%. Kemudian, peubah penting yang dihasilkan dari pemodelan *Random Forest* adalah besar daya listrik, penghasilan ayah, dan jumlah anak dalam keluarga.

Kata kunci: *adaboost*, klasifikasi, *random forest*, uang kuliah tunggal

@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.

ABSTRACT

FAREL FIRMAN. Modelling of Late Payment Tuition Fee Classification of Students IPB University using Random Forest and AdaBoost. Supervised by MOHAMMAD MASJKUR dan CICI SUHAENI.

According to the data obtained from Bureau of Finance IPB University, there was an increase the number of students who experienced delays in paying tuition fee in each generation starting in 2016 until 2018 with number of students who experienced delays in paying tuition fee are 54, 64, and 117 students. This study aims to compare classification modelling using Random Forest and AdaBoost. Furthermore, this study also identify important variables that are obtained from the best model. Random Forest final classifier using optimal parameter number of explanatory variables as many as 7 variables and number of trees as many as 500 trees with Random Oversampling (ROS) imbalanced data method. AdaBoost final classifier using optimal parameter iteration as many as 80 iterations with Random Undersampling (RUS) imbalanced data method. Random Forest and AdaBoost classifier have performance metric ROC-AUC that are 58,70% and 52,90%, respectively. Random Forest classifier has better prediction performance than AdaBoost classifier on predict student who will late payment tuition fee with a difference in value of ROC-AUC is 5,8%. Then, important variables obtained from best classifier Random Forest are amount of electric power, father's income, and the number of children in the family.

Keywords: adaboost, classification, random forest, tuition fee

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 IPBUniversity.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPBUniversity.



PEMODELAN KLASIFIKASI KETERLAMBATAN PEMBAYARAN UKT MAHASISWA IPB DENGAN RANDOM FOREST DAN ADABOOST

FAREL FIRMAN

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
2021**

- 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 IPBUniversity.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPBUniversity.



@Hak cipta milik IPBUniversity

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 IPBUniversity.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPBUniversity.

Tim Penguji pada Ujian Skripsi:

1. Gerry Alfa Dito, M.Si.



Judul Skripsi: **Pemodelan Klasifikasi Keterlambatan Pembayaran UKT**
Mahasiswa IPB dengan *Random Forest* dan *AdaBoost*

Nama : Farel Firman
NIM : G14160046

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:
Ir. Mohammad Masjkur, M.S.



Pembimbing 2:
Cici Suhaeni, M.Si.



Diketahui oleh

Ketua Departemen:
Dr. Anang Kurnia, M.Si.
NIP. 197308241997021001



Tanggal Ujian:
13 Januari 2021

Tanggal Lulus: 04 FEB 2021

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 IPBUniversity

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 IPBUniversity.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPBUniversity.



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah *subhanahu wa ta'ala* yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan karya ilmiah ini dengan judul *Pemodelan Klasifikasi Keterlambatan Pembayaran UKT Mahasiswa IPB dengan Random Forest dan AdaBoost*.

Terselesainya karya ilmiah ini tidak lepas dari berbagai pihak yang telah membantu penulis dalam kelancaran proses penulisan, oleh karena itu penulis ingin mengucapkan terima kasih kepada:

1. Bapak Ir. Mohammad Masjkur, M.S dan Ibu Cici Suhaeni, M.Si selaku komisi pembimbing yang telah banyak memberikan saran, bimbingan, dan waktunya untuk menyelesaikan karya ilmiah ini.
2. Bapak Gerry Alfa Dito, M.Si selaku dosen penguji luar yang telah memberikan saran dan koreksi penulisan karya ilmiah ini.
3. Seluruh dosen dan staff Departemen Statistika IPB yang telah memberikan ilmu yang bermanfaat serta menunjang segala kebutuhan administrasi penulis.
4. Ayah, Ibu, Farid, Fadea, serta seluruh keluarga atas doa dan kasih sayang yang senantiasa diberikan kepada penulis.
5. Teman-teman seperjuangan Statistika 53 yang telah memberi dukungan kepada penulis selama perkuliahan hingga terselesainya karya ilmiah ini.

Semoga Allah *subhanahu wa ta'ala* membalas semua kebaikan yang telah mereka berikan kepada penulis dan semoga karya ilmiah ini dapat bermanfaat dalam bidang Statistika.

Bogor, Februari 2021

Farel Firman



@Hak cipta milik IPBUniversity

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 IPBUniversity.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPBUniversity.



DAFTAR ISI

| | |
|--|----|
| DAFTAR TABEL | vi |
| DAFTAR GAMBAR | vi |
| PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Tujuan Penelitian | 1 |
| TINJAUAN PUSTAKA | 2 |
| 2.1 <i>Classification and Regression Tree (CART)</i> | 2 |
| 2.2 Random Forest | 2 |
| 2.3 AdaBoost | 3 |
| 2.4 Ketidakseimbangan Data | 4 |
| 2.5 <i>K-Fold Stratified Cross Validation</i> | 6 |
| 2.6 Ukuran Kinerja Model Klasifikasi | 6 |
| III METODE | 8 |
| 3.1 Data | 8 |
| 3.2 Prosedur Analisis Data | 9 |
| IV HASIL DAN PEMBAHASAN | 11 |
| 4.1 Gambaran Umum Karakteristik Mahasiswa | 11 |
| 4.2 Pemodelan Klasifikasi Random Forest | 14 |
| 4.2.1 Pemodelan Klasifikasi Random Forest dengan RUS | 14 |
| 4.2.2 Pemodelan Klasifikasi Random Forest dengan ROS | 15 |
| 4.2.3 Pemodelan Klasifikasi Random Forest dengan SMOTE | 15 |
| 4.2.4 Penentuan Nilai Parameter Optimal pada Random Forest | 16 |
| 4.3 Pemodelan Klasifikasi dengan AdaBoost | 16 |
| 4.3.1 Pemodelan AdaBoost dengan RUS | 16 |
| 4.3.2 Pemodelan AdaBoost dengan ROS | 17 |
| 4.3.3 Pemodelan AdaBoost dengan SMOTE | 17 |
| 4.3.4 Penentuan Nilai Parameter Optimal pada AdaBoost | 18 |
| 4.4 Perbandingan Random Forest dan AdaBoost | 18 |
| V SIMPULAN DAN SARAN | 21 |
| 5.1 Simpulan | 21 |
| 5.2 Saran | 21 |
| DAFTAR PUSTAKA | 23 |
| RIWAYAT HIDUP | 24 |

Hak cipta milik IPB University

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.

DAFTAR TABEL

| | | |
|---|---|----|
| 1 | Matriks konfusi | 6 |
| 2 | Daftar peubah yang digunakan | 8 |
| 3 | Nilai parameter optimal Random Forest pada masing-masing pendekatan <i>resampling</i> | 16 |
| 4 | Nilai parameter optimal AdaBoost pada masing-masing pendekatan <i>resampling</i> | 18 |

DAFTAR GAMBAR

| | | |
|----|---|----|
| 1 | Proporsi mahasiswa yang mengalami keterlambatan pembayaran UKT | 11 |
| 2 | Persentase jenis kelamin dan asal daerah mahasiswa terhadap keterlambatan pembayaran UKT | 11 |
| 3 | Persentase jalur masuk penerimaan dan besar daya listrik terhadap keterlambatan pembayaran UKT | 12 |
| 4 | Sebaran jumlah anak dan persentase urutan anak dalam keluarga terhadap keterlambatan pembayaran UKT | 12 |
| 5 | Persentase pendidikan ayah dan pendidikan ibu terhadap keterlambatan pembayaran UKT | 13 |
| 6 | Persentase penghasilan ayah dan penghasilan ibu terhadap keterlambatan pembayaran UKT | 13 |
| 7 | Hasil ROC-AUC terhadap banyak pohon (k) berdasarkan banyak peubah penjelas (m) menggunakan <i>Random Undersampling</i> | 14 |
| 8 | Hasil ROC-AUC terhadap banyak pohon (k) berdasarkan banyak peubah penjelas (m) menggunakan <i>Random Oversampling</i> | 15 |
| 9 | Hasil ROC-AUC terhadap banyak pohon (k) berdasarkan banyak peubah penjelas (m) menggunakan <i>Synthetic Minority Oversampling Technique</i> | 16 |
| 10 | Hasil ROC-AUC terhadap banyak iterasi (h) menggunakan <i>Random Undersampling</i> | 17 |
| 11 | Hasil ROC-AUC terhadap banyak iterasi (h) menggunakan <i>Random Oversampling</i> | 17 |
| 12 | Hasil ROC-AUC terhadap banyak iterasi (h) menggunakan <i>Synthetic Minority Oversampling Technique</i> | 18 |
| 13 | Peubah penting Random Forest | 19 |
| 14 | Peubah penting AdaBoost | 20 |
| 15 | Perbandingan Kurva ROC Random Forest dan AdaBoost | 20 |

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.