



PERBANDINGAN MODEL REGRESI LOGISTIK BINER DENGAN DAN TANPA SMOTE PADA KETEPATAN WAKTU KELULUSAN MAHASISWA PROGRAM SARJANA IPB

DHEA PUSPITA ADINDA



**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.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Perbandingan Model Regresi Logistik Biner dengan dan Tanpa *SMOTE* pada Ketepatan Waktu Kelulusan Mahasiswa Program Sarjana IPB” 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, Juli 2024

Dhea Puspita Adinda
G1401201090

- 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

DHEA PUSPITA ADINDA. Perbandingan Model Regresi Logistik Biner dengan dan Tanpa *SMOTE* pada Ketepatan Waktu Kelulusan Mahasiswa Program Sarjana IPB. Dibimbing oleh RAHMA ANISA dan UTAMI DYAH SYAFITRI.

IPB memiliki tujuan menghasilkan lulusan unggul. Hal ini dapat dicapai dengan menghasilkan mahasiswa lulus tepat waktu. Mahasiswa program sarjana dianggap lulus tepat waktu jika menyelesaikan studi dalam waktu maksimal empat tahun. Oleh karena itu, penting dilakukan penelitian untuk mengetahui pengaruh faktor-faktor terhadap ketepatan waktu kelulusan mahasiswa. Salah satu analisis yang digunakan untuk melakukan analisis terhadap respon yang biner (lulus tepat waktu/tidak) adalah analisis regresi logistik biner. Data yang digunakan adalah data wisudawan program sarjana IPB bulan Agustus dan September 2023. Proporsi lulusan tidak tepat waktu sebesar 30% menunjukkan data relatif tidak seimbang, sehingga diterapkan metode *SMOTE*. Penelitian ini bertujuan untuk membandingkan performa model dengan dan tanpa *SMOTE* serta mengetahui besar pengaruh faktor-faktor terhadap ketepatan waktu kelulusan mahasiswa. Hasil yang diperoleh menunjukkan model regresi logistik biner tanpa *SMOTE* lebih baik dibandingkan model dengan *SMOTE* berdasarkan nilai sensitivitas model tanpa *SMOTE* yang lebih besar dibandingkan model dengan *SMOTE* yaitu sebesar 90,7%. Hal ini disebabkan ketidakseimbangan data didominasi oleh mahasiswa lulus tepat waktu, sehingga model tanpa *SMOTE* mampu menggambarkan mahasiswa lulus tepat waktu dengan baik. Model menunjukkan bahwa perbedaan fakultas, IPK, predikat kelulusan, beasiswa, jenis kelamin, jalur masuk, SKS, UKT, dan keikutsertaan organisasi memiliki pengaruh yang berbeda terhadap kecenderungan mahasiswa lulus tepat waktu.

Kata kunci: data tidak seimbang, ketepatan waktu kelulusan, regresi logistik biner, *SMOTE*

ABSTRACT

DHEA PUSPITA ADINDA. Comparison of Binary Logistic Regression with and without SMOTE on the Timeliness of IPB Undergraduate Students. Supervised by RAHMA ANISA and UTAMI DYAH SYAFITRI.

IPB aims to produce outstanding graduates by ensuring students graduate on time. Undergraduate students are considered to graduate on time if they complete their program within four years. Therefore, it is important to research factors influencing the punctuality of student graduation. Binary logistic regression analysis is used to analyze the binary response. This research used IPB undergraduate program graduates dataset in August and September 2023. The percentage of graduates who did not graduate on time is 30%, implying relatively unbalanced class in the dataset, thus the SMOTE method was applied. This research aims to compare the performance of models with and without SMOTE and to determine the factors influencing the punctuality of student graduation. The results showed that the binary logistic regression model without SMOTE was better than the model with SMOTE. Based on the sensitivity, which was higher at 90,7% compared to the model with SMOTE. Due to the imbalance in data dominated by students graduating on time, the model without SMOTE accurately represents students graduating on time. The model showed that differences in faculty, GPA, graduation predicate, scholarship, gender, admission path, cumulative credit, tuition fee, and organizational participation have different effects on the timeliness of student graduation.

Keywords: binary logistic regression analysis, graduate on-time, SMOTE, unbalanced data



© 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.

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.

PERBANDINGAN MODEL REGRESI LOGISTIK BINER DENGAN DAN TANPA SMOTE PADA KETEPATAN WAKTU KELULUSAN MAHASISWA PROGRAM SARJANA IPB

DHEA PUSPITA ADINDA

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 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.

Tim Penguji pada Ujian Skripsi:
Akbar Rizki, S.Stat., 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 : Perbandingan Model Regresi Logistik Biner dengan dan Tanpa SMOTE pada Ketepatan Waktu Kelulusan Mahasiswa Program Sarjana IPB

Nama : Dhea Puspita Adinda
NIM : G1401201090

@Hak cipta milik IPB University

Disetujui oleh

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



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



Diketahui oleh

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



Tanggal Ujian:
11 Juli 2024

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 2024 sampai bulan Juli 2024 ini ialah ketepatan waktu kelulusan mahasiswa, dengan judul “Perbandingan Model Regresi Logistik Biner dengan dan Tanpa SMOTE pada Ketepatan Waktu Kelulusan Mahasiswa Program Sarjana IPB”.

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

1. Ibu Rahma Anisa, S.Stat., M.Si. dan Ibu Dr. Utami Dyah Syafitri, S.Si., M.Si. yang telah membimbing dan banyak memberi saran selama penyusunan karya ilmiah ini;
2. Ibu Akbar Rizki, S.Stat., M.Si selaku dosen penguji skripsi yang telah memberikan banyak saran dan masukan dalam penulisan karya ilmiah ini.;
3. Seluruh dosen dan tenaga pendidik Departemen Statistika IPB yang telah memberikan ilmu dan membantu kebutuhan administrasi penulis;
4. Direktorat Administrasi Pendidikan IPB yang telah memberikan data yang digunakan dalam penelitian ini;
5. Ibu, bapak, kakak, dan seluruh keluarga yang telah memberikan dukungan, doa, dan kasih sayangnya kepada penulis;
6. Sahabat dan teman-teman yang selalu kebersamai dalam penulisan karya ilmiah ini khususnya Mayunda, Bunga, Salma, Annisa, Alisa, Naura, Arsy, Nanda, Tias, Tsabitah, Berliana, Hanaa, Hanung, Raffael, dan seluruh teman-teman statistika 57; dan
7. Seluruh pihak yang tidak dapat penulis sebutkan satu persatu atas segala kebaikan, doa, dukungan, dan motivasi hingga penulis dapat menyelesaikan karya ilmiah ini.

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

Bogor, Juli 2024

Dhea Puspita Adinda



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xi
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan Penelitian	1
TINJAUAN PUSTAKA	2
2.1 Analisis Regresi Logistik Biner	2
2.2 Multikolinearitas	4
2.3 <i>Syntetic Minority Oversampling Technique (SMOTE)</i>	4
2.4 <i>K-fold Cross Validation</i>	5
2.5 Ketepatan Klasifikasi	6
III METODE	8
3.1 Data	8
3.2 Prosedur Analisis Data	9
IV HASIL DAN PEMBAHASAN	11
4.1 Eksplorasi Data	11
4.2 Pengecekan Multikolinearitas	13
4.3 Analisis Regresi Logistik Tanpa Penerapan SMOTE	14
4.4 Analisis Regresi Logistik dengan Penerapan SMOTE	15
4.5 Pemilihan Model Terbaik	16
4.6 Interpretasi Odds Ratio	18
4.5 Penerapan Model Regresi Logistik Biner	20
V SIMPULAN DAN SARAN	21
5.1 Simpulan	21
5.2 Saran	21
DAFTAR PUSTAKA	22
LAMPIRAN	24
RIWAYAT HIDUP	28

Hak cipta milik 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.

DAFTAR TABEL

1	Tabel klasifikasi	6
2	Peubah yang digunakan dalam penelitian	8
3	Hasil pemeriksaan multikolinearitas	14
4	Hasil uji parsial model tanpa SMOTE	14
5	Performa model tanpa SMOTE setiap <i>fold</i>	15
6	Hasil uji parsial model dengan SMOTE	15
7	Performa model dengan SMOTE setiap <i>fold</i>	16
8	Perbandingan performa model berdasarkan <i>k-fold cross validation</i>	17
9	Hasil klasifikasi model tanpa SMOTE	17
10	Nilai rasio odds	18

DAFTAR GAMBAR

1	Ilustrasi <i>10-fold cross validation</i>	6
2	Persentase ketepatan waktu kelulusan wisudawan	11
3	Persentase ketepatan waktu kelulusan wisudawan berdasarkan fakultas	11
4	Persentase ketepatan waktu kelulusan wisudawan berdasarkan predikat kelulusan	12
5	Persentase ketepatan waktu kelulusan wisudawan berdasarkan IPK	12
6	Persentase ketepatan waktu kelulusan wisudawan berdasarkan jenis kelamin	12
7	Persentase ketepatan waktu kelulusan wisudawan berdasarkan UKT	13
8	Persentase ketepatan waktu kelulusan wisudawan berdasarkan keterlibatan dalam organisasi	13

DAFTAR LAMPIRAN

1	Proporsi ketepatan waktu kelulusan wisudawan berdasarkan karakteristik wisudawan	25
2	Distribusi wisudawan berdasarkan ketepatan waktu kelulusan	26
3	Hasil pengujian parameter	26



@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.