

**IDENTIFIKASI PEUBAH YANG BERPENGARUH TERHADAP  
KETIDAKLULUSAN MAHASISWA PROGRAM SARJANA  
BUD IPB DENGAN REGRESI LOGISTIK BINER**

**MAHDIYAH RIAESNIANDA**



**DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2022**

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

Dengan ini saya menyatakan bahwa skripsi dengan judul “Identifikasi Peubah yang Berpengaruh terhadap Ketidاكلulusan Mahasiswa Program Sarjana BUD IPB dengan Regresi Logistik Biner” 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, September 2022

Mahdiah Riaesnianda  
G14180062

## ABSTRAK

MAHDIYAH RIAESNIANDA. Identifikasi Peubah yang Berpengaruh terhadap Ketidاكلulusan Mahasiswa Program Sarjana BUD IPB dengan Regresi Logistik Biner. Dibimbing oleh AAM ALAMUDI dan AGUS M. SOLEH.

Salah satu jalur masuk yang ada di Institut Pertanian Bogor (IPB) yaitu Beasiswa Utusan Daerah (BUD). Tidak semua mahasiswa BUD IPB dapat menyelesaikan studinya karena di *Drop Out* (DO) atau mengundurkan diri. Salah satu upaya yang dapat dilakukan IPB untuk menekan angka putus kuliah mahasiswa BUD IPB adalah dengan mengetahui peubah yang berpengaruh terhadap ketidاكلulusan mahasiswa BUD IPB. Peubah yang berpengaruh terhadap ketidاكلulusan mahasiswa BUD IPB dianalisis dengan regresi logistik biner. Terdapat ketidakseimbangan kelas data pada peubah respon sehingga metode yang dapat digunakan untuk mengatasi hal tersebut adalah *Synthetic Minority Over-Sampling Technique* (SMOTE). Model klasifikasi dengan SMOTE menghasilkan rata-rata sensitivitas yang lebih tinggi dibandingkan model tanpa SMOTE dari 10,66% menjadi 61,91%. Hal tersebut memperkuat bahwa model dengan SMOTE lebih baik untuk memprediksi kelas minoritas (mahasiswa BUD IPB yang tidak lulus). Peubah yang berpengaruh terhadap ketidاكلulusan mahasiswa BUD IPB adalah jenis kelamin, status sekolah asal, kelompok program studi, ada tidaknya Program Pra Universitas (PPU), jenis sponsor, rata-rata rapor, dan IPK Tahap Persiapan Bersama (TPB) atau Program Pendidikan Kompetensi Umum (PPKU).

Kata kunci: BUD IPB, regresi logistik biner, SMOTE

## ABSTRACT

MAHDIYAH RIAESNIANDA. Identification of Variables Affecting the Failure of BUD IPB Undergraduate Students with Binary Logistics Regression. Supervised by AAM ALAMUDI and AGUS M. SOLEH.

One of the entrances available at the Bogor Agricultural University (IPB) is the Regional Representatives Scholarship (BUD). Not all BUD IPB students were able to complete their studies because they *dropped out* (DO) or resigned. One of the efforts that IPB can do to reduce the dropout rate for BUD IPB students is to find out the variables that affect the failure of BUD IPB students. The variables that influence the failure of BUD IPB students are analyzed by binary logistic regression. There is an imbalance of data classes in the response variables so that the method that can be used to overcome this is the Synthetic Minority Over-Sampling Technique (SMOTE). The classification model with SMOTE resulted in a higher average sensitivity than the model without SMOTE from 10,66% to 61,91%. This confirms that the model with SMOTE is better at predicting the minority class (BUD IPB students who do not pass). The variables that affect the failure of BUD IPB students are gender, school status of origin, study program groups, the presence or absence of Pre-University Programs (PPU), type of sponsor, average report cards, and GPA in the Joint Preparation Stage (TPB) or General Competency Education Program (PPKU).

Keywords: binary logistic regression, BUD IPB, SMOTE

© Hak Cipta milik IPB, tahun 2022  
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 YANG BERPENGARUH TERHADAP  
KETIDAKLULUSAN MAHASISWA PROGRAM SARJANA  
BUD IPB DENGAN REGRESI LOGISTIK BINER**

**MAHDIYAH RIAESNIANDA**

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  
2022**

Tim Penguji pada Ujian Skripsi:

1 Dra. Itasia Dina Sulvianti, M.Si.

Judul Skripsi : Identifikasi Peubah yang Berpengaruh terhadap Ketidaklulusan  
Mahasiswa Program Sarjana BUD IPB dengan Regresi Logistik  
Biner

Nama : Mahdiyah Riaesnianda  
NIM : G14180062

Disetujui oleh

Pembimbing 1:

Ir. Aam Alamudi, M.Si.

Pembimbing 2:

Dr. Agus M. Soleh, S.Si., M.T.

Diketahui oleh

Ketua Departemen:

Dr. Anang Kurnia, M.Si.

NIP. 197308241997021001

Tanggal Ujian: 25 Agustus 2022

Tanggal Lulus:

## **PRAKATA**

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga penulis dapat menyelesaikan karya ilmiahnya dengan judul “Identifikasi Peubah yang Berpengaruh terhadap Ketidaklulusan Mahasiswa Program Sarjana BUD IPB dengan Regresi Logistik Biner”.

Karya ilmiah ini dapat diselesaikan karena penulis mendapatkan banyak doa, dukungan, dan saran. Terima kasih penulis ucapkan kepada:

1. Para pembimbing, Bapak Ir. Aam Alamudi, M.Si. dan Bapak Dr. Agus M. Soleh, S.Si., M.T. yang telah membimbing dan banyak memberi saran kepada penulis serta Ibu Dra. Itasia Dina Sulvianti, M.Si. selaku dosen penguji pada ujian sidang penulis.
2. Ayahanda Aria Purnama, ibunda Eko Sari Kurniasih, saudara kandung Ares Gusti Nugraha, dan keluarga yang selalu memberi dukungan dan doa kepada penulis.
3. Seluruh dosen Departemen Statistika IPB yang telah membagikan ilmu dan pengalamannya kepada penulis selama masa perkuliahan.
4. Staf Tata Usaha Departemen Statistika IPB yang telah membantu dalam hal administrasi.
5. Teman-teman Departemen Statistika angkatan 55 yang telah kebersamai penulis selama masa perkuliahan.
6. Sahabat-sahabat penulis yang selalu memberikan dukungan dan perhatian selama penulis menyelesaikan karya ilmiah ini.
7. Semua pihak yang telah membantu dan memberi dukungan kepada penulis.

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

Bogor, September 2022

*Mahdiyah Riaesnianda*



## DAFTAR ISI

DAFTAR TABEL	viii
DAFTAR GAMBAR	viii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 Beasiswa Utusan Daerah (BUD) IPB	3
2.2 Multikolinearitas	3
2.3 Regresi Logistik Biner	4
2.4 Evaluasi Model Klasifikasi	7
2.5 <i>Synthetic Minority Over-Sampling Technique</i> (SMOTE)	8
III METODE	9
3.1 Data	9
3.2 Prosedur Analisis Data	9
IV HASIL DAN PEMBAHASAN	12
4.1 Eksplorasi Data	12
4.2 Regresi Logistik Biner	16
4.3 Regresi Logistik Biner dengan SMOTE	17
4.4 Perbandingan Kinerja Klasifikasi Regresi Logistik Biner	18
4.5 Peubah-peubah yang Berpengaruh terhadap Ketidاكلulusan Mahasiswa BUD IPB	20
V SIMPULAN	23
DAFTAR PUSTAKA	24
RIWAYAT HIDUP	26

## DAFTAR TABEL

1	Matriks konfusi	7
2	Performa klasifikasi model dengan AUC	7
3	Daftar peubah	9
4	Hasil pemeriksaan multikolinearitas	16
5	Jumlah data sebelum SMOTE	16
6	Jumlah data setelah SMOTE	17
7	Perbandingan rata-rata kinerja klasifikasi regresi logistik biner	19
8	Hasil uji Wald regresi logistik biner	20
9	Nilai rasio odds	21

## DAFTAR GAMBAR

1	Persentase status kelulusan mahasiswa BUD IPB	12
2	<i>Boxplot</i> peubah penjas (a) rata-rata rapor (b) IPK TPB/PPKU	12
3	Persentase status kelulusan berdasarkan jenis kelamin	13
4	Persentase status kelulusan berdasarkan status sekolah asal	13
5	Persentase status kelulusan berdasarkan akreditasi sekolah	14
6	Persentase status kelulusan berdasarkan kelompok program studi	14
7	Persentase status kelulusan berdasarkan PPU	15
8	Persentase status kelulusan berdasarkan tipe sponsor	15
9	Kinerja klasifikasi model regresi logistik tanpa SMOTE	17
10	Kinerja klasifikasi model regresi logistik dengan SMOTE	18
11	Perbandingan (a) akurasi (b) sensitivitas (c) spesifisitas (d) AUC	19

# I PENDAHULUAN

## 1.1 Latar Belakang

Institut Pertanian Bogor (IPB) hingga saat ini telah membuka program kuliah sarjana dengan beberapa jalur penerimaan, yaitu Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN), Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN), Ujian Tulis Mandiri Berbasis Komputer (UTM-BK), Prestasi Internasional Nasional (PIN), Ketua Osis, dan Beasiswa Utusan Daerah (BUD) (Admisi IPB 2021). Jalur BUD IPB merupakan jalur masuk kelembagaan. Jalur ini mengharuskan mahasiswa yang ingin mendaftarkan diri memiliki sponsor atau penyanggah dana. Selama masa perkuliahan hingga waktu yang disepakati antara mitra dan IPB, biaya pendidikan mahasiswa BUD IPB didanai oleh pihak penyanggah dana atau sponsor.

Ketatnya proses seleksi dan persyaratan untuk menjadi calon mahasiswa BUD IPB tidak menjamin mahasiswa yang berhasil diterima dapat menyelesaikan studinya untuk meraih gelar sarjana. Hal tersebut dibuktikan dengan tidak semua mahasiswa BUD IPB berhasil lulus. Mahasiswa BUD IPB yang putus kuliah terdiri atas mahasiswa yang di *Drop Out* (DO) dan mengundurkan diri. Data yang diperoleh dari Direktorat Administrasi Pendidikan dan Penerimaan Mahasiswa Baru (DAPPMB) IPB menunjukkan angka putus kuliah pada jalur masuk BUD IPB tahun 2012 hingga 2017 secara berturut-turut adalah 1,9%, 2,8%, 3,6%, 3,9%, 5,7%, 9,0% dari total mahasiswa BUD IPB setiap tahunnya. Meningkatnya angka putus kuliah dapat menjadi evaluasi bagi IPB dalam membuat suatu program atau kebijakan agar angka putus kuliah mahasiswa BUD IPB tidak meningkat pada masa mendatang. Angka putus kuliah disebut baik jika terjadi penurunan karena menunjukkan mutu pendidikan. Salah satu upaya yang dapat dilakukan IPB untuk menekan angka putus kuliah mahasiswa BUD IPB yaitu dengan mengetahui peubah yang berpengaruh terhadap ketidakkulusan mahasiswa BUD IPB.

Penelitian yang berkaitan dengan peubah yang berpengaruh terhadap putus kuliah dan keberhasilan studi mahasiswa program sarjana pernah dilakukan oleh beberapa peneliti terdahulu. Menurut Imran *et al.* (2013), peubah yang berhubungan dengan mahasiswa program sarjana reguler IPB angkatan 2008 yang mengalami putus kuliah adalah jenis kelamin, IPK, dan fakultas. Penelitian lain oleh Rahmah (2013), menyatakan bahwa peubah yang berpengaruh terhadap keberhasilan studi mahasiswa BUD IPB angkatan 45 hingga 48 adalah pulau asal, nilai rata-rata ujian nasional, sponsor, penghasilan orang tua, dan pendidikan ibu.

Salah satu indikator untuk melihat keberhasilan studi mahasiswa BUD IPB yaitu status lulus dan tidak lulus mahasiswa setelah melaksanakan proses perkuliahan. Peubah-peubah yang berpengaruh terhadap ketidakkulusan mahasiswa BUD IPB dapat dianalisis dengan regresi logistik biner. Penelitian dengan menggunakan analisis regresi logistik biner pernah dilakukan oleh Nurganita (2015) dalam mengidentifikasi peubah-peubah yang memengaruhi ketepatan waktu kelulusan mahasiswa program sarjana IPB. Regresi logistik biner memungkinkan peubah penjelas bertipe numerik dan kategorik untuk memprediksi peubah respon bertipe kategorik (Osborne 2007).

Peubah respon yang digunakan pada penelitian ini adalah status lulus dan tidak lulus. Terdapat ketidakseimbangan kelas data pada peubah respon yaitu

mahasiswa BUD IPB yang lulus lebih banyak dibandingkan mahasiswa BUD IPB yang tidak lulus. Permasalahan ketidakseimbangan kelas data ini perlu ditangani karena dapat menyebabkan model yang diprediksi cenderung pada kelas mayoritas sehingga mengakibatkan tingginya kesalahan klasifikasi pada kelas minoritas (Chawla *et al.* 2002). Salah satu metode yang dapat digunakan untuk mengatasi hal ini adalah *Synthetic Minority Over-Sampling Technique* (SMOTE). Penelitian terdahulu oleh Amelia *et al.* (2018) menggunakan SMOTE dalam penanganan ketidakseimbangan data dapat meningkatkan kinerja *Support Vector Machine* (SVM) dalam mengklasifikasikan mahasiswa yang tidak lulus.

Penelitian mengenai regresi logistik biner dan SMOTE pernah dilakukan oleh Pangestika *et al.* (2021) yang menerapkan SMOTE pada pemodelan regresi logistik biner terhadap keberhasilan studi mahasiswa program magister IPB. Hasil penelitiannya menunjukkan bahwa model hasil penerapan SMOTE dengan *oversampling* 600% meningkatkan ketepatan klasifikasi (spesifisitas) kelas minoritas (mahasiswa yang tidak lulus) sebesar 30,55%. Penelitian lain oleh Widasari (2017) yang menerapkan metode SMOTE pada model regresi logistik biner data pekerja tahun 2014 di Kabupaten Badung, menyatakan bahwa model dengan SMOTE *oversampling* 300% mampu meningkatkan ketepatan klasifikasi (sensitivitas) pada kelas minoritas (pekerja sektor informal) sebesar 45,45%.

## 1.2 Tujuan

Tujuan penelitian ini adalah mengetahui peubah yang berpengaruh terhadap ketidاكلulusan mahasiswa BUD IPB.

## II TINJAUAN PUSTAKA

### 2.1 Beasiswa Utusan Daerah (BUD) IPB

BUD IPB merupakan salah satu jalur penerimaan program sarjana IPB. Jalur BUD IPB pertama kali dibuka pada tahun 2004. Jalur ini menjaring putra-putri berprestasi dari berbagai daerah yang tidak terjangkau dengan jalur reguler lainnya. Setiap calon mahasiswa BUD IPB direkomendasikan dan dibiayai oleh pemerintah pusat, pemerintah daerah, perusahaan atau lembaga swasta (Admisi IPB 2021). Jalur ini dibuka dengan harapan putra-putri bangsa lulusan jalur BUD IPB tersebut dapat membangun daerahnya menjadi lebih baik. Kriteria yang dibutuhkan calon mahasiswa BUD IPB, yaitu berumur tidak lebih dari 25 tahun, memiliki nilai rapor SMA lima semester pertama mata pelajaran (Matematika, Fisika, Kimia, Biologi) minimal 7 (rentang 0 – 10) atau 70 (rentang 0 – 100), sehat jasmani dan rohani, bebas narkoba, serta bersedia tinggal di asrama Tingkat Persiapan Bersama (TPB) / Program Pendidikan Kompetensi Umum (PPKU) IPB pada tahun pertama (BUD IPB 2021).

Penyandang dana atau pihak sponsor memiliki peraturan yang berbeda-beda untuk setiap mahasiswa yang dibiayainya. Salah satunya terdapat pihak sponsor yang menerapkan Program Pra Universitas (PPU). Program ini merupakan suatu bentuk persiapan sebelum mahasiswa BUD IPB memasuki masa TPB/PPKU dengan harapan mahasiswa BUD IPB dapat melalui proses perkuliahan dengan baik. PPU dapat berlangsung selama satu tahun atau dapat lebih singkat seperti dua bulan. Perbedaan durasi PPU berdasarkan kesepakatan antara mitra dan IPB.

### 2.2 Multikolinearitas

Multikolinearitas adalah suatu keadaan adanya hubungan yang kuat antara dua peubah penjelas atau lebih dalam model regresi. Menurut Garson (2008), asumsi yang harus dipenuhi dalam regresi logistik adalah tidak adanya multikolinearitas pada peubah penjelas. Multikolinearitas akan mengakibatkan galat baku dari koefisien regresinya membesar dan pada pengujian parameter secara simultan, *p-value* yang dihasilkan signifikan namun ketika pengujian parameter secara parsial tidak terdapat peubah penjelas yang signifikan (Agresti 2007). Multikolinearitas dapat dideteksi dengan Korelasi *Pearson*. Permasalahan multikolinearitas terjadi apabila nilai Korelasi *Pearson* diantara dua peubah penjelas lebih besar dari 0,8 atau 0,9 (Senaviratna dan Cooray 2019).

$$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)}\sqrt{(\sum y_i^2 - n\bar{y}^2)}} \quad (1)$$

Keterangan:

$r$  = koefisien korelasi

$n$  = banyaknya amatan

$x_i$  = nilai ke- $i$  peubah  $x$

$y_i$  = nilai ke- $i$  peubah  $y$

$\bar{x}$  = rata-rata dari peubah  $x$

$\bar{y}$  = rata-rata dari peubah  $y$

Selain menggunakan nilai Korelasi *Pearson*, multikolinearitas dapat dideteksi dengan nilai *Variance Inflation Factor* (VIF). Multikolinearitas diindikasikan berpengaruh kuat apabila nilai  $VIF > 10$  yang sesuai dengan nilai koefisien Korelasi *Pearson* 0,95. Namun, permasalahan multikolinearitas dengan dampak yang serius mungkin terjadi pada nilai VIF yang lebih rendah seperti 3 hingga 5 sehingga diperlukan pemeriksaan kembali terkait pengaruh dari multikolinearitas tersebut (Hair *et al.* 2010).

$$VIF = \frac{1}{1 - R^2} \quad (2)$$

dengan  $R^2$  adalah koefisien determinasi dari peubah penjelas  $x$  dengan peubah penjelas lainnya. Menurut Fox dan Monette (1992), VIF tidak dapat secara langsung digunakan jika terdapat peubah kategorik dengan lebih dari satu derajat bebas sehingga disarankan menggunakan *Generalized Variance Inflation Factor* ( $GVIF^{\frac{1}{2 \times db}}$ ).  $GVIF^{\frac{1}{2 \times db}}$  merupakan VIF yang terkoreksi oleh derajat bebas (db) dari peubah penjelas. Nilai  $GVIF^{\frac{1}{2 \times db}}$  yang dikuadratkan setara dengan nilai VIF untuk peubah kontinu.

$$VIF = \left( GVIF^{\frac{1}{2 \times db}} \right)^2 \quad (3)$$

### 2.3 Regresi Logistik Biner

Analisis regresi merupakan suatu analisis statistika yang digunakan untuk memperoleh model serta mendeskripsikan hubungan antara peubah respon dan satu atau lebih peubah penjelas. Regresi logistik biner merupakan bagian dari analisis regresi dengan bentuk peubah respon yang bersifat dikotomik atau biner (Hosmer *et al.* 2013). Menurut Hosmer *et al.* (2013), model regresi logistik secara umum:

$$\pi(\mathbf{x}) = \left[ \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \right] \quad (4)$$

Keterangan:

$\pi(\mathbf{x})$  =  $P(Y = 1|\mathbf{x}) = 1 - P(Y = 0|\mathbf{x})$  dan  $0 \leq \pi(\mathbf{x}) \leq 1$

$\beta_0$  = konstanta

$\beta_i$  = koefisien regresi logistik

$x_i$  = peubah penjelas

$i$  =  $1, \dots, p$

$p$  = banyaknya peubah penjelas

Model regresi tersebut bersifat *non linear* sehingga untuk mengubahnya menjadi linear diperlukan transformasi yaitu transformasi logit:

$$\text{logit} [\pi(\mathbf{x})] = \ln \left[ \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = g(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (5)$$

Menurut Hosmer *et al.* (2013), apabila peubah penjelas berskala nominal memiliki  $k$  kategori maka diperlukan  $k-1$  peubah boneka (*dummy*) dalam model. Sehingga transformasi logit menjadi:

$$g(\mathbf{x}) = \beta_0 + \sum_{j=1}^p \beta_j x_j + \sum_{l=1}^{k_j-1} \beta_{jl} D_{jl} \quad (6)$$

Keterangan:

$\beta_0$  = konstanta

$\beta_j, \dots, \beta_p$  = koefisien regresi logistik

$x_j, \dots, x_p$  = peubah penjelas

$\beta_{jl}$  = koefisien peubah boneka pada peubah penjelas ke- $j$  dan kategori ke- $l$

$D_{jl}$  = peubah boneka pada peubah penjelas ke- $j$  dan kategori ke- $l$

$k_j$  = banyaknya kategori pada peubah penjelas ke- $j$

Metode kemungkinan maksimum merupakan suatu pendekatan untuk menduga parameter dalam suatu model regresi logistik. Fungsi kemungkinan maksimum apabila antar amatan diasumsikan saling bebas:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} (1 - \pi(\mathbf{x}_i))^{1-y_i} \quad (7)$$

Keterangan:

$n$  = banyaknya amatan

$i$  =  $1, 2, \dots, n$

$y_i$  = respon pada amatan ke- $i$

$\pi(\mathbf{x}_i)$  = peluang kejadian amatan ke- $i$  bernilai  $y = 1$

Pendugaan parameter  $\boldsymbol{\beta}$  dilakukan dengan memaksimumkan *log-likelihood* sebagai berikut:

$$L(\boldsymbol{\beta}) = \ln[l(\boldsymbol{\beta})] = \sum_{i=1}^n \{y_i \ln[\pi(\mathbf{x}_i)] + (1 - y_i) \ln[1 - \pi(\mathbf{x}_i)]\} \quad (8)$$

Nilai  $\hat{\boldsymbol{\beta}}$  diperoleh dengan menurunkan  $L(\boldsymbol{\beta})$  terhadap  $\beta_i = 0$  secara iteratif dengan metode Newton-Raphson (Jennrich dan Robinson 1969).

Setelah memperoleh dugaan parameter model regresi logistik, langkah selanjutnya adalah melakukan uji signifikansi parameter untuk mengetahui apakah peubah penjelas dalam model secara nyata berpengaruh terhadap peubah respon. Menurut Hosmer *et al.* (2013), menguji parameter secara simultan atau keseluruhan menggunakan statistik uji G.

Hipotesisnya sebagai berikut:

$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$  (tidak terdapat pengaruh peubah penjelas secara simultan terhadap peubah respon),

$H_1: \exists \beta_i \neq 0$ , dengan  $i = 1, 2, \dots, p$  (terdapat paling sedikit satu peubah penjelas yang berpengaruh terhadap peubah respon)

Statistik uji untuk uji G:

$$G = -2 \ln \frac{L_0}{L_p} \sim \chi^2_{(db)} \quad (9)$$

dengan  $L_0$  adalah fungsi kemungkinan maksimum tanpa peubah penjelas dan  $L_p$  adalah fungsi kemungkinan maksimum dengan peubah penjelas. Apabila hasil uji G menunjukkan terdapat paling sedikit satu peubah penjelas yang berpengaruh terhadap peubah respon, langkah selanjutnya adalah menguji parameter secara parsial.

Menurut Hosmer *et al.* (2013), menguji parameter secara parsial menggunakan statistik uji *Wald*.

Hipotesisnya sebagai berikut:

$H_0: \beta_i = 0$  dengan  $i = 1, 2, \dots, p$  (peubah penjelas ke- $i$  tidak berpengaruh terhadap peubah respon)

$H_1: \beta_i \neq 0$ , dengan  $i = 1, 2, \dots, p$  (peubah penjelas ke- $i$  berpengaruh terhadap peubah respon)

Statistik uji untuk uji *Wald*:

$$W_i = \left[ \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \right]^2 \sim \chi^2_{(db)} \quad (10)$$

dengan  $\hat{\beta}_i$  adalah nilai dugaan untuk parameter  $\beta_i$ ,  $SE(\hat{\beta}_i)$  adalah galat baku untuk  $\hat{\beta}_i$ .

Rasio Odds merupakan suatu ukuran asosiasi yang biasa digunakan untuk menginterpretasikan koefisien dalam regresi logistik. Rasio odds merupakan perbandingan antara odds dari dua kategori. Odds merupakan peluang kejadian sukses terjadi terhadap peluang kejadian sukses tersebut tidak terjadi. Menurut Agresti (2007), odds didefinisikan dengan persamaan sebagai berikut.

$$Odds = \frac{\pi_i}{1 - \pi_i} \quad (11)$$

dengan  $\pi_i$  merupakan peluang kejadian sukses ( $Y=1$ ) dan  $1 - \pi_i$  merupakan peluang kejadian gagal ( $Y=0$ ). Menurut Hosmer *et al.* (2013), rasio odds dilambangkan dengan OR. Hubungan rasio odds dengan koefisien regresi adalah sebagai berikut.

$$OR = \exp(\beta_i) \quad (12)$$



## 2.4 Evaluasi Model Klasifikasi

Menurut Chawla *et al.* (2002), kinerja suatu sistem klasifikasi biasanya dievaluasi dengan matriks konfusi. Matriks konfusi terdiri dari kolom yang memuat kelas prediksi dan baris yang memuat kelas sebenarnya, seperti pada Tabel 1.

Tabel 1 Matriks konfusi

Kelas Aktual	Kelas Prediksi		Ketepatan Klasifikasi
	Positif	Negatif	
Positif	TP	FN	Sensitivitas = $\frac{TP}{TP+FN}$
Negatif	FP	TN	Spesifisitas = $\frac{TN}{TN+FP}$
			Akurasi = $\frac{TP+TN}{TP+TN+FN+FP}$

Keterangan:

*True Positive* (TP) : Jumlah data positif yang terklasifikasi dengan benar oleh sistem

*True Negative* (TN) : Jumlah data negatif yang terklasifikasi dengan benar oleh sistem

*False Negative* (FN) : Jumlah data positif yang terklasifikasi negatif oleh sistem

*False Positive* (FP) : Jumlah data negatif yang terklasifikasi positif oleh sistem

Berbagai kriteria performa bagi suatu sistem klasifikasi seperti, akurasi, spesifisitas, sensitivitas, dan lain-lain dapat diturunkan dari matriks konfusi (Pramana *et al.* 2018). Akurasi merupakan pengukuran performa suatu sistem klasifikasi yang paling umum digunakan. Mengevaluasi performa suatu sistem klasifikasi pada *imbalanced* data tidak dapat menggunakan nilai akurasi karena akurasi menempatkan bobot yang lebih besar pada kelas mayoritas. Menurut Provost dan Fawcett (2001), *Receiver Operating Characteristics* (ROC) dapat menilai kinerja klasifikasi pada *imbalanced* data. ROC merupakan kurva dua dimensi dengan sumbu vertikal *True Positive Rate* (TPR) dan sumbu horizontal *False Positive Rate* (FPR). Menurut Hosmer *et al.* (2013), area di bawah kurva ROC (Area Under Curve (AUC)) dapat mendeskripsikan akurasi suatu klasifikasi dengan lebih baik dan lebih lengkap. AUC berkisar antara 0 hingga 1 yang menunjukkan keseluruhan pengukuran atas kesesuaian dari model yang digunakan. Secara umum, acuan pengukuran kebaikan model dalam membedakan kelas positif dan kelas negatif dengan nilai AUC disajikan pada Tabel 2.

Tabel 2 Performa klasifikasi model dengan AUC

Nilai AUC	Kategori Model
0,5	model tidak disarankan
0,5 < AUC < 0,7	model dianggap buruk
0,7 ≤ AUC < 0,8	model dianggap dapat diterima
0,8 ≤ AUC < 0,9	model dianggap baik
AUC ≥ 0,9	model dianggap sangat baik

## 2.5 Synthetic Minority Over-Sampling Technique (SMOTE)

*Synthetic Minority Over-Sampling Technique* (SMOTE) merupakan modifikasi dari metode *oversampling* dengan membangkitkan data baru (data sintetis) pada kelas minoritas berdasarkan *k-nearest neighbor* sehingga antar kelas data mayoritas dan kelas minoritas menjadi seimbang (Chawla *et al.* 2002). SMOTE yang digunakan ketika data terdiri atas peubah penjelas bertipe numerik dan kategorik adalah SMOTE-NC (*Nominal Continuous*). Penentuan jarak tetangga terdekat dihitung dengan jarak *Euclidean* dan menambahkan nilai median dari simpangan baku semua peubah bertipe numerik untuk kelas minoritas. Nilai median dimasukkan dalam perhitungan jarak *Euclidean* jika peubah kategorik antara amatan dan tetangga terdekatnya berbeda. Persamaan jarak *Euclidean* adalah sebagai berikut.

$$d(\mathbf{a}, \mathbf{b}) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (13)$$

Keterangan:

$d(\mathbf{a}, \mathbf{b})$  = jarak *Euclidean* untuk dua vektor amatan

$a_i$  = nilai amatan  $\mathbf{a}$  untuk peubah penjelas ke- $i$

$b_i$  = nilai amatan  $\mathbf{b}$  untuk peubah penjelas ke- $i$

$n$  = banyaknya peubah penjelas

$i$  = 1, 2, ...,  $n$

Setelah menghitung jarak tetangga terdekat, data baru dibangkitkan dengan metode SMOTE. Data numerik dibangkitkan dengan persamaaan:

$$\mathbf{x}_{baru} = \mathbf{x} + (\mathbf{x}^* - \mathbf{x}) \times rand [0,1] \quad (14)$$

dengan  $\mathbf{x}_{baru}$  adalah vektor amatan baru,  $\mathbf{x}$  adalah vektor amatan awal,  $\mathbf{x}^*$  adalah vektor amatan *k-nearest neighbor*, dan  $rand [0,1]$  adalah bilangan acak antara 0 dan 1. Sementara untuk data kategorik, kategori yang paling sering muncul antara amatan dan *k-nearest neighbor* dijadikan sebagai amatan baru.

### III METODE

#### 3.1 Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Direktorat Administrasi Pendidikan dan Penerimaan Mahasiswa Baru (DAPMB) IPB. Data tersebut merupakan data mahasiswa BUD IPB tahun masuk 2012 – 2017 yang sudah memiliki status kelulusan. Data terdiri atas 925 amatan dengan 878 mahasiswa BUD IPB berstatus lulus dan 47 mahasiswa BUD IPB berstatus tidak lulus. Penelitian ini lebih fokus terhadap model dalam mengklasifikasikan mahasiswa BUD IPB yang tidak lulus. Peubah yang digunakan pada penelitian ini terdiri atas satu peubah respon dan delapan peubah penjelas. Peubah respon adalah status kelulusan mahasiswa BUD IPB, yaitu lulus ( $Y = 0$ ) dan tidak lulus ( $Y = 1$ ). Terdapat enam peubah penjelas bertipe kategorik dan dua peubah penjelas bertipe numerik. Daftar peubah yang digunakan disajikan pada Tabel 3.

Tabel 3 Daftar peubah

Peubah	Nama Peubah	Keterangan
$Y$	Status Kelulusan	Lulus = 0 Tidak Lulus = 1
$X_1$	Jenis Kelamin	Laki-Laki = 0 Perempuan = 1
$X_2$	Status SLTA	Negeri = 0 Swasta = 1
$X_3$	Akreditasi	BT = 0 C = 1 B = 2 A = 3
$X_4$	Kelompok Program Studi	Sains dan Teknologi = 0 Sains dan Sosial = 1
$X_5$	Program Pra Universitas (PPU) BUD IPB	Tidak ada = 0 Ada = 1
$X_6$	Tipe Sponsor	Perusahaan = 0 Pemerintah kabupaten/kota = 1 Pemerintah provinsi = 2 Pemerintah pusat = 3 Yayasan/lembaga lainnya = 4
$X_7$	Rataan Rapor	-
$X_8$	IPK TPB/PPKU	-

#### 3.2 Prosedur Analisis Data

Penelitian ini dianalisis menggunakan R 4.0.2. Tahapan-tahapan analisis yang dilakukan sebagai berikut:

- Melakukan pra-proses data. Langkah-langkah yang dilakukan:

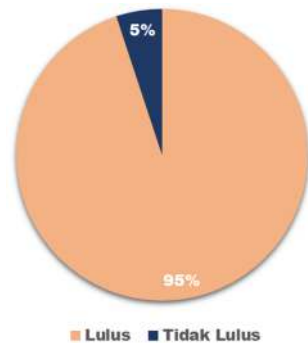
- 1) Melakukan seleksi terhadap mahasiswa BUD IPB yang memiliki data tidak lengkap.
  - 2) Memetakan peubah akreditasi yang merujuk pada Badan Akreditasi Nasional Sekolah/Madrasah (BAN-S/M).
  - 3) Mengelompokkan peubah jenis sponsor merujuk pada *website* BUD IPB.
  - 4) Mengelompokkan peubah program studi menjadi dua, yaitu kelompok program studi Sains dan Teknologi (ST) yang terdiri atas semua program studi di Fakultas Pertanian, Fakultas Kedokteran Hewan, Fakultas Perikanan dan Ilmu Kelautan, Fakultas Peternakan, Fakultas Kehutanan dan Lingkungan, Fakultas Teknologi Pertanian, Fakultas MIPA, dan Program Studi Ilmu Gizi. Sementara kelompok Sains dan Sosial (SS) terdiri atas semua program studi di Fakultas Ekonomi dan Manajemen, Program Studi Ilmu Keluarga dan Konsumen, Program Studi Komunikasi dan Pengembangan Masyarakat, dan Program Studi Bisnis.
- b. Melakukan eksplorasi data untuk mengetahui gambaran umum mahasiswa BUD IPB berdasarkan status kelulusannya. Langkah-langkah yang dilakukan:
- 1) Membuat diagram lingkaran untuk mengetahui persentase mahasiswa BUD IPB yang lulus dan tidak lulus.
  - 2) Membuat *boxplot* pada peubah numerik berdasarkan kategori mahasiswa BUD IPB yang lulus dan tidak lulus.
  - 3) Membuat diagram batang pada peubah kategorik berdasarkan kategori mahasiswa BUD IPB yang lulus dan tidak lulus.
- c. Membagi data secara acak menjadi 80% data latih dan 20% data uji .
- d. Melakukan analisis regresi logistik biner dengan tahapan sebagai berikut.
- 1) Membangun model dengan data latih.
  - 2) Melakukan evaluasi model dengan *confusion matrix* dan nilai AUC dari data uji.
- e. Melakukan SMOTE dengan tahapan sebagai berikut.
- 1) Menghitung median dari standar deviasi semua peubah numerik di kelas minoritas. Jika peubah kategorik berbeda diantara amatan terpilih dan tetangga terdekatnya, maka nilai median dimasukkan dalam perhitungan jarak *Euclidean*.
  - 2) Menghitung jarak antar amatan pada kelas minoritas menggunakan persamaan jarak *Euclidean*.
  - 3) Menentukan nilai k yaitu tiga untuk k-tetangga terdekat.
  - 4) Satu contoh amatan dipilih secara acak dari kelas minoritas.
  - 5) Menentukan amatan k-tetangga terdekat dengan mengurut jarak contoh amatan terpilih dengan semua amatan pada kelas minoritas.
  - 6) Membangkitkan data buatan berdasarkan amatan k tetangga terdekat.
  - 7) Langkah empat sampai enam dilakukan hingga jumlah kelas minoritas dan kelas mayoritas menjadi seimbang.
- f. Membangun model dari data latih yang telah melalui tahap SMOTE dengan regresi logistik biner dan melakukan evaluasi model dengan *confusion matrix* serta nilai AUC dari data uji.
- g. Mengulangi tahap c hingga f sebanyak 100 kali.

- h. Membandingkan hasil evaluasi kedua model klasifikasi (tanpa tahap SMOTE dan melalui tahap SMOTE) dengan melihat *boxplot* dari *confusion matrix* dan nilai AUC.
- i. Membangun model dengan data keseluruhan berdasarkan hasil perbandingan evaluasi model pada tahap h.
- j. Melakukan pengujian signifikansi parameter secara simultan dan parsial.
- k. Interpretasi model berdasarkan nilai rasio *odds* dari peubah-peubah yang berpengaruh signifikan.

## IV HASIL DAN PEMBAHASAN

### 4.1 Eksplorasi Data

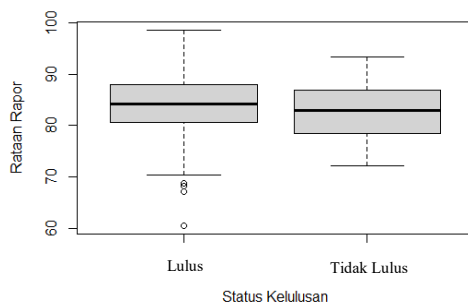
Eksplorasi data dilakukan untuk melihat gambaran umum mahasiswa BUD IPB berdasarkan status kelulusannya. Data yang digunakan sebanyak 925 mahasiswa BUD IPB yang sudah memiliki status kelulusan. Persentase status kelulusan mahasiswa BUD IPB dapat dilihat pada Gambar 1.



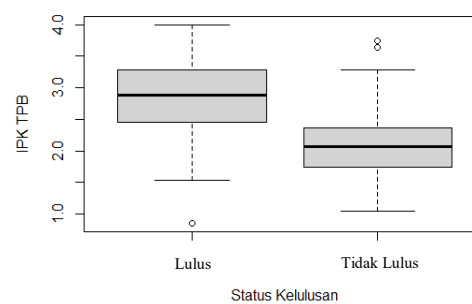
Gambar 1 Persentase status kelulusan mahasiswa BUD IPB

Gambar 1 menunjukkan jumlah mahasiswa BUD IPB yang berstatus lulus lebih banyak dibandingkan mahasiswa BUD IPB yang berstatus tidak lulus. Mahasiswa BUD IPB yang berstatus lulus memiliki persentase sebesar 95% sementara mahasiswa BUD IPB yang berstatus tidak lulus memiliki persentase sebesar 5%. Perbedaan persentase yang jauh tersebut mengindikasikan adanya ketidakseimbangan data antara kelas lulus dan tidak lulus.

Peubah penjelas numerik yang digunakan adalah rata-rata rapor dan IPK TPB/PPKU. Gambar 2(a) menunjukkan mayoritas mahasiswa BUD IPB yang lulus memiliki rata-rata rapor 80,68 hingga 87,94. Beberapa mahasiswa BUD IPB yang berstatus lulus memiliki nilai rata-rata rapor di bawah 70. Beberapa mahasiswa tersebut menjalani tes tambahan berupa Tes Potensi Akademik (TPA) pada saat seleksi awal masuk IPB. Rata-rata rapor maksimum mahasiswa BUD IPB yang berstatus lulus yaitu 98,60 dan mahasiswa BUD IPB yang berstatus tidak lulus yaitu 93,33.



(a)

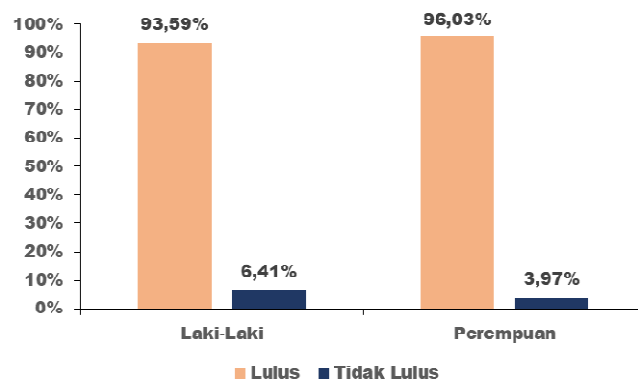


(b)

Gambar 2 *Boxplot* peubah penjelas (a) rata-rata rapor (b) IPK TPB/PPKU

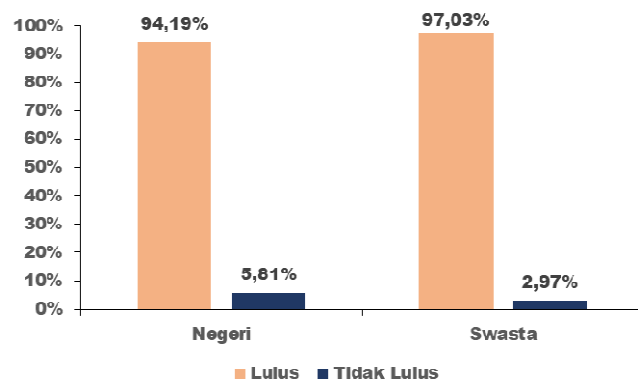
Gambar 2(b) menunjukkan median IPK TPB/PPKU mahasiswa BUD IPB yang berstatus lulus lebih besar dibandingkan mahasiswa BUD IPB yang tidak lulus. Mahasiswa BUD IPB yang berstatus lulus memiliki nilai IPK TPB/PPKU yang lebih beragam dibandingkan mahasiswa BUD IPB yang tidak lulus. Terdapat beberapa mahasiswa BUD IPB dengan status lulus memiliki nilai IPK TPB/PPKU kurang dari sama dengan 1,70. Mahasiswa tersebut dengan tahun masuk 2012 hingga 2013 yang memiliki ketetapan untuk nilai IPK TPB/PPKU kurang dari sama dengan 1,50 status kelanjutan studinya adalah dikeluarkan sementara untuk mahasiswa dengan nilai IPK TPK/PPKU dibawah 1,00 perlu ditelusuri lebih lanjut. Beberapa mahasiswa BUD IPB dengan status tidak lulus memiliki nilai IPK TPB/PPKU di atas 3,50.

Persentase status kelulusan mahasiswa BUD IPB berdasarkan jenis kelamin dapat dilihat pada Gambar 3. Mahasiswa BUD IPB dengan jenis kelamin perempuan memiliki persentase 54,49% lebih banyak dibandingkan laki-laki. Akan tetapi, Mahasiswa BUD IPB yang berjenis kelamin laki-laki dengan status tidak lulus lebih banyak sebesar 6,41% dibandingkan perempuan sebesar 3,97%.



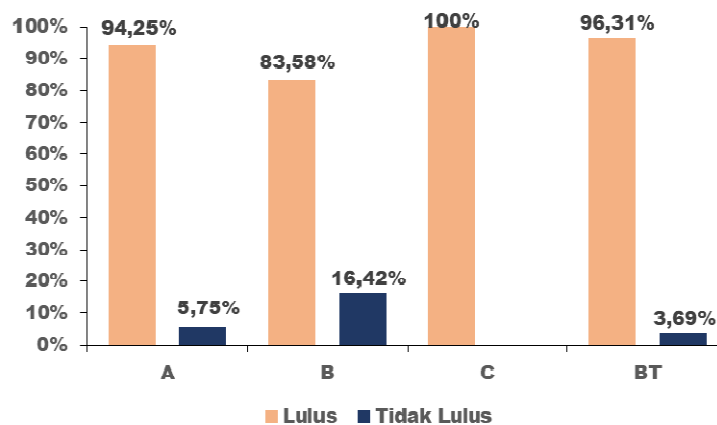
Gambar 3 Persentase status kelulusan berdasarkan jenis kelamin

Gambar 4 menunjukkan persentase status kelulusan berdasarkan status sekolah asal. Mahasiswa BUD IPB yang berasal dari sekolah negeri memiliki persentase sebesar 74,49% lebih banyak dibandingkan sekolah swasta. Mahasiswa BUD IPB yang tidak lulus dengan sekolah asal berstatus negeri sebesar 5,81% lebih banyak dibandingkan mahasiswa BUD IPB dengan sekolah asal swasta (2,97%).



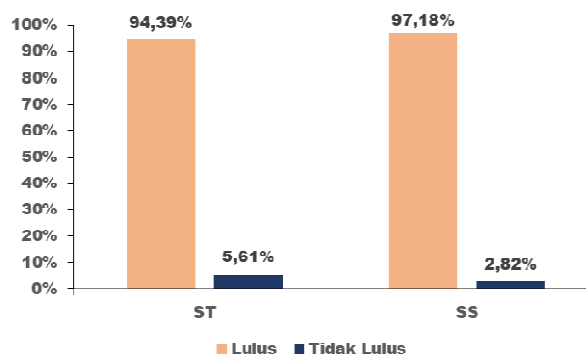
Gambar 4 Persentase status kelulusan berdasarkan status sekolah asal

Persentase status kelulusan mahasiswa BUD IPB berdasarkan akreditasi sekolah asal terlihat pada Gambar 5. Mahasiswa BUD IPB yang berasal dari sekolah Belum Terakreditasi (BT) persentasenya lebih banyak dibandingkan sekolah yang berasal dari akreditasi lainnya, yaitu sebesar 67,35%. Mahasiswa BUD IPB yang berstatus tidak lulus dengan akreditasi sekolah asal yaitu B memiliki persentase sebesar 16,42% lebih banyak dibandingkan akreditasi lainnya. Semua mahasiswa BUD IPB yang berasal dari sekolah dengan akreditasi C berstatus lulus.



Gambar 5 Persentase status kelulusan berdasarkan akreditasi sekolah

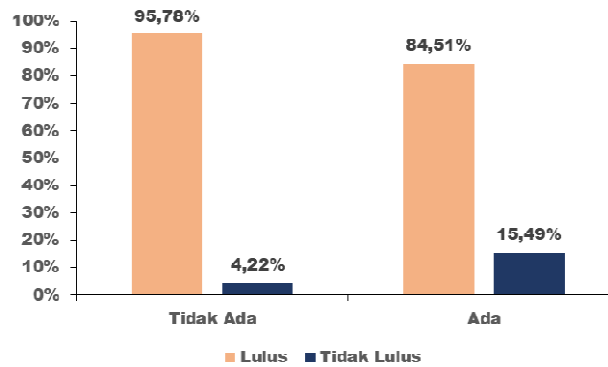
Gambar 6 menunjukkan persentase status kelulusan berdasarkan kelompok program studi. Mahasiswa BUD IPB dengan kelompok program studi Sains dan Teknologi memiliki persentase yang lebih banyak yaitu 80,86%. Mahasiswa BUD IPB dengan status tidak lulus dari kelompok program studi Sains dan Teknologi (ST) memiliki persentase 5,61% lebih banyak dibandingkan kelompok program studi Sains dan Sosial (SS) (2,82%).



Gambar 6 Persentase status kelulusan berdasarkan kelompok program studi

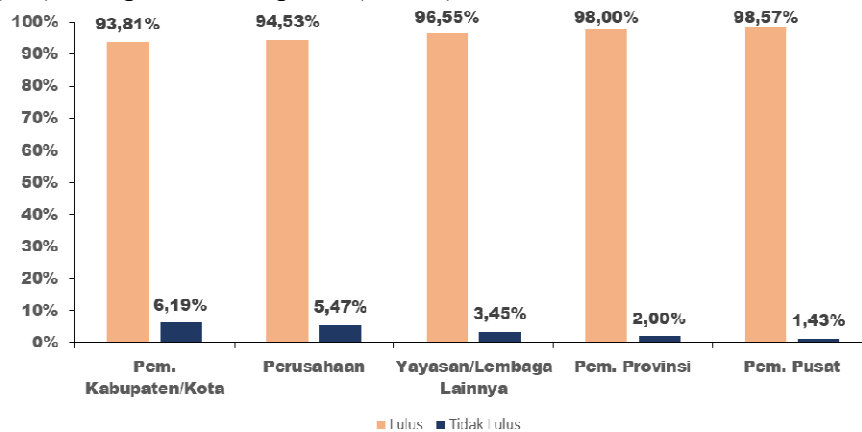
Persentase status kelulusan berdasarkan PPU BUD IPB ditunjukkan pada Gambar 7. Mahasiswa BUD IPB yang tidak mengikuti PPU lebih banyak sebesar 92,32%. Akan tetapi, mahasiswa BUD IPB dengan status tidak lulus lebih banyak pada mahasiswa BUD IPB yang mengikuti PPU, yaitu sebesar 15,49% berbeda jauh dengan mahasiswa BUD IPB yang tidak mengikuti PPU (4,22%).





Gambar 7 Persentase status kelulusan berdasarkan PPU

Gambar 8 menunjukkan persentase status kelulusan berdasarkan tipe sponsor. Persentase mahasiswa BUD IPB dengan tipe sponsor pemerintah kabupaten/kota adalah 48,86% lebih banyak dibandingkan tipe sponsor yang lainnya. Persentase mahasiswa BUD IPB dengan status tidak lulus untuk tipe sponsor pemerintah kabupaten/kota sebesar 6,19% lebih banyak dibandingkan tipe sponsor perusahaan (5,47%), tipe sponsor yayasan/lembaga lainnya (3,45%), tipe sponsor pemerintah provinsi (2%), dan pemerintah pusat (1,43%).



Gambar 8 Persentase status kelulusan berdasarkan tipe sponsor

Langkah pertama sebelum dilakukan analisis regresi logistik biner yaitu pemeriksaan apakah terdapat masalah multikolinearitas yang serius antar peubah penjelas atau tidak. Nilai Korelasi *Pearson* antara peubah IPK TPB/PPKU dan nilai rata-rapor sebesar 0,367 yang mengindikasikan hubungan yang lemah. Hal ini menunjukkan tidak adanya masalah serius terkait multikolinearitas antara peubah penjelas numerik.

Selain menggunakan Korelasi *Pearson*, multikolinearitas dapat dideteksi dengan nilai VIF. Seperti menurut Fox dan Monette (1992), pada penelitian ini digunakan peubah bertipe kategorik dengan beberapa kategori sehingga nilai VIF tidak dapat secara langsung digunakan untuk melihat ada atau tidaknya permasalahan multikolinearitas pada peubah penjelasnya.  $G\text{VIF} \frac{1}{2 \times db}$  yang dikuadratkan digunakan untuk memeriksa multikolinearitas pada peubah penjelas karena  $\left(G\text{VIF} \frac{1}{2 \times db}\right)^2$  akan setara dengan aturan VIF pada peubah kontinu.

Tabel 4 menunjukkan nilai  $\left(GVIF \frac{1}{2 \times db}\right)^2$  untuk setiap peubah penjelas mendekati nilai satu. Hal ini tidak mengindikasikan adanya permasalahan multikolinearitas yang serius sehingga proses analisis regresi logistik biner dapat dilanjutkan.

Tabel 4 Hasil pemeriksaan multikolinearitas

Peubah	db	$GVIF \frac{1}{2 \times db}$	$\left(GVIF \frac{1}{2 \times db}\right)^2$
$X_1$	1	1,0242	1,0490
$X_2$	1	1,0867	1,1809
$X_3$	3	1,0474	1,0970
$X_4$	1	1,0443	1,0906
$X_5$	1	1,1503	1,3232
$X_6$	4	1,0723	1,1498
$X_7$	1	1,0980	1,2056
$X_8$	1	1,1093	1,2305

## 4.2 Regresi Logistik Biner

Sebelum dilakukan analisis regresi logistik biner, data dibagi secara acak menjadi 80% untuk data latih data uji sebesar 20% dengan kategori lulus sebesar 95% dan kategori tidak lulus sebesar 5%. Penelitian yang dilakukan oleh Musu *et al.* (2021) menunjukkan bahwa persentase komposisi data 80:20 pada tiga data *set* yang berbeda menghasilkan akurasi yang tinggi dibandingkan komposisi data 50:50, 60:40, dan 70:30. Pembagian data latih dan data uji dilakukan dengan pengulangan sebanyak seratus kali. Pengulangan dilakukan untuk mengurangi ketidakpastian pendugaan dari performa sistem klasifikasi. Banyaknya data masing-masing kelas lulus dan tidak lulus tanpa SMOTE terlihat pada Tabel 5.

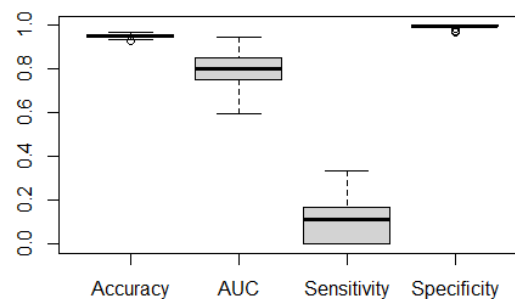
Tabel 5 Banyaknya data sebelum SMOTE

Kelas Data	Data Latih	Data Uji
Lulus	703	175
Tidak Lulus	38	9
Total	741	184

Model regresi logistik biner dibangun dari setiap *set* data latih. Kinerja model dalam mengklasifikasikan mahasiswa BUD IPB yang lulus dan tidak lulus dengan benar diprediksi menggunakan data uji pada masing-masing pengulangan. Kebaikan model dalam mengklasifikasikan data dengan benar diukur dengan akurasi, sensitivitas, spesifisitas, dan AUC seluruh pengulangan yang disajikan pada Gambar 9.

Gambar 9 menunjukkan akurasi, spesifisitas dan AUC yang dihasilkan tinggi sementara sensitivitas yang dihasilkan model rendah. Rataan akurasi yang dihasilkan dari keseluruhan evaluasi model sebesar 95,20% yang berarti 95,20% mahasiswa BUD IPB benar diklasifikasikan lulus dan tidak lulus (sesuai dengan status kelulusannya). Rataan sensitivitas yang dihasilkan evaluasi model seluruh pengulangan yaitu 10,66% yang berarti 10,66% mahasiswa BUD IPB tepat

diklasifikasikan tidak lulus. Rataan spesifisitas menunjukkan 99,52% mahasiswa BUD IPB tepat diklasifikasikan lulus. Rataan nilai AUC yang dihasilkan dari seluruh pengulangan evaluasi model adalah 0,7994 yang berarti kebaikan model dalam membedakan mahasiswa BUD IPB yang lulus dan tidak lulus sebesar 0,7994. Nilai AUC tersebut menunjukkan model tanpa SMOTE dapat diterima (Hosmer *et al.* 2013). Akan tetapi, sensitivitas pada keseluruhan model klasifikasi sangat rendah. Hal ini menunjukkan bahwa model cenderung memprediksi kelas mayoritas (mahasiswa BUD IPB yang lulus) sehingga menyebabkan kesalahan klasifikasi pada kelas minoritas (mahasiswa BUD IPB yang tidak lulus). Model dianggap belum dapat mengklasifikasikan mahasiswa BUD IPB yang tidak lulus dengan baik. Hal ini dapat terjadi karena banyaknya amatan mahasiswa BUD IPB pada kelas yang lulus lebih banyak dibandingkan mahasiswa BUD IPB yang tidak lulus (data tidak seimbang) sehingga diperlukan penanganan agar hasil kinerja klasifikasi menjadi lebih baik.



Gambar 9 Kinerja klasifikasi model regresi logistik tanpa SMOTE

### 4.3 Regresi Logistik Biner dengan SMOTE

SMOTE dilakukan dengan membangkitkan data baru dengan menerapkan *k-nearest neighbours* untuk setiap data di kelas minoritas. Pembangkitan data dilakukan hingga banyaknya amatan mahasiswa BUD IPB pada kelas yang lulus sama dengan mahasiswa BUD IPB yang tidak lulus. Persentase banyaknya amatan mahasiswa BUD IPB yang tidak lulus sebelum dilakukan SMOTE sebesar 5% dari total amatan. Setelah diterapkan SMOTE pada kelas minoritas, persentase banyaknya amatan mahasiswa BUD IPB yang tidak lulus menjadi 50% dari total amatan. Banyaknya amatan pada masing-masing kelas lulus dan tidak lulus dengan SMOTE terlihat pada Tabel 6.

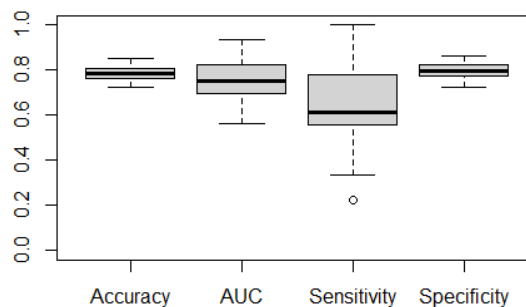
Tabel 6 Banyaknya data setelah SMOTE

Kelas Data	Data Latih	Data Latih setelah SMOTE
Lulus	703	703
Tidak Lulus	38	703
Total	741	1.406

Model regresi logistik dibangun dari data latih awal dengan tambahan data buatan hasil SMOTE. Kinerja model dalam mengklasifikasikan mahasiswa BUD IPB sesuai status kelulusannya diprediksi menggunakan data uji yang sama dengan

sebelum SMOTE pada masing-masing pengulangan. Hasil evaluasi model dari seluruh pengulangan dengan penerapan SMOTE disajikan pada Gambar 10.

Gambar 10 menunjukkan sensitivitas yang dihasilkan model dengan SMOTE meningkat dibandingkan model tanpa SMOTE. Rataan akurasi yang dihasilkan dari keseluruhan evaluasi model menurun menjadi 78,81% yang berarti 78,81% mahasiswa BUD IPB benar diklasifikasi lulus dan tidak lulus (sesuai dengan status kelulusannya). Rataan sensitivitas yang dihasilkan dari evaluasi model seluruh pengulangan meningkat yaitu menjadi 61,91% yang berarti 61,91% mahasiswa BUD IPB tepat diklasifikasikan tidak lulus. Rataan spesifisitas menunjukkan penurunan menjadi 79,68% yang berarti 79,68% mahasiswa BUD IPB tepat diklasifikasikan lulus. Rataan nilai AUC yang dihasilkan dari seluruh pengulangan evaluasi model menurun menjadi 0,7574 yang berarti kebaikan model dalam membedakan mahasiswa BUD IPB yang lulus dan tidak lulus sebesar 0,7574. Nilai AUC tersebut menunjukkan model dengan SMOTE masih dalam kategori dapat diterima (Hosmer *et al.* 2013). Menurunnya akurasi, spesifisitas, dan AUC terjadi karena adanya peningkatan mahasiswa BUD IPB yang lulus diklasifikasikan salah sebagai mahasiswa BUD IPB yang tidak lulus. Hal ini menunjukkan bahwa model sudah tidak cenderung memprediksi mahasiswa BUD IPB yang lulus. Model regresi logistik biner dengan SMOTE dianggap sudah dapat mengklasifikasikan mahasiswa BUD IPB yang tidak lulus dengan lebih baik dilihat dari rataan sensitivitas yang meningkat dan menurunnya rataan akurasi, spesifisitas, dan nilai AUC.



Gambar 10 Kinerja klasifikasi model regresi logistik dengan SMOTE

#### 4.4 Perbandingan Kinerja Klasifikasi Regresi Logistik Biner

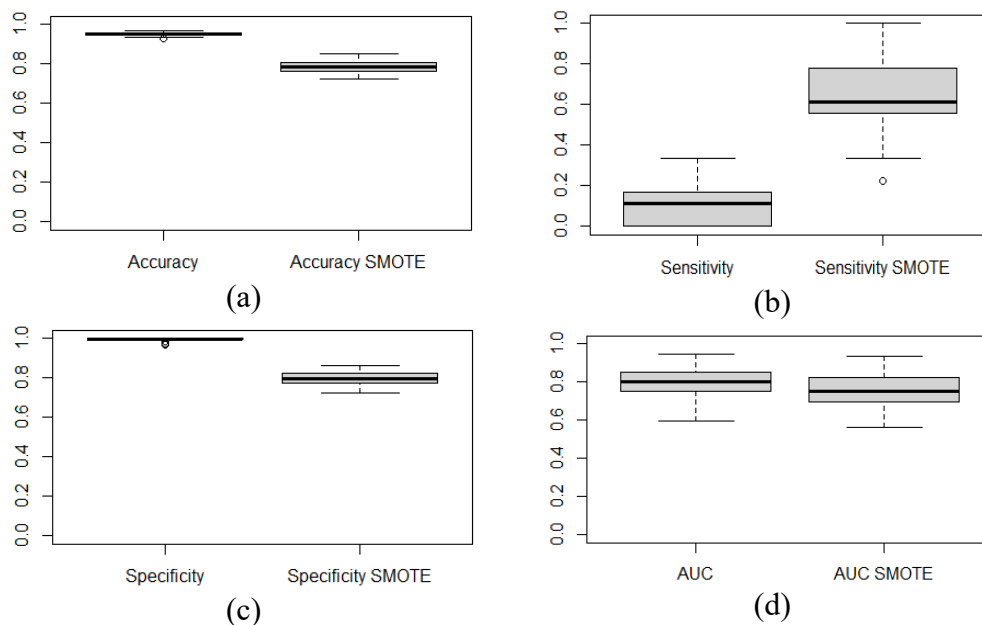
Hasil kinerja klasifikasi model sebelum SMOTE dan setelah SMOTE dibandingkan untuk menentukan model yang terbaik untuk mengklasifikasikan mahasiswa BUD IPB yang tidak lulus. Perbandingan rata-rata akurasi, sensitivitas, spesifisitas, dan nilai AUC dari kedua model disajikan pada Tabel 7.

Tabel 7 menunjukkan model dengan SMOTE menghasilkan rata-rata sensitivitas yang lebih tinggi dibandingkan model tanpa SMOTE. Hal tersebut memperkuat bahwa model dengan SMOTE mampu mengklasifikasikan mahasiswa BUD IPB yang tidak lulus dengan baik karena kesalahan mahasiswa BUD IPB yang tidak lulus diklasifikasikan sebagai mahasiswa BUD IPB yang lulus rendah. Gambar 11 menunjukkan perbandingan akurasi, sensitivitas, spesifisitas, dan AUC dari model regresi logistik biner tanpa SMOTE dan dengan SMOTE.

Tabel 7 Perbandingan rata-rata kinerja klasifikasi regresi logistik biner

Model	Rata-Rata Kinerja Klasifikasi			
	Akurasi	Sensitivitas	Spesifisitas	AUC
Tanpa SMOTE	95,20%	10,66%	99,52%	0,7994
Dengan SMOTE	78,81%	61,91%	79,68%	0,7574

Gambar 11 menunjukkan akurasi, sensitivitas, spesifisitas, dan AUC dari model dengan SMOTE lebih beragam dibandingkan model tanpa SMOTE. Akurasi yang dihasilkan model dengan SMOTE cenderung menyebar pada rentang 76,60% hingga 81,00% lebih rendah jika dibandingkan dengan model tanpa SMOTE yaitu pada rentang 94,60% hingga 95,70%. Sensitivitas yang dihasilkan model dengan SMOTE cenderung menyebar pada rentang 55,60% hingga 77,80%. Persentase tersebut lebih tinggi jika dibandingkan model tanpa SMOTE yang cenderung menyebar pada rentang 0% hingga 13,88%. Spesifisitas yang dihasilkan model dengan SMOTE cenderung menyebar pada rentang 77,55% hingga 82,30% lebih rendah jika dibandingkan dengan model tanpa SMOTE yaitu 99,40% hingga 100%. Nilai AUC yang dihasilkan model dengan SMOTE cenderung menyebar pada rentang 0,6955 hingga 0,8260. Nilai tersebut lebih rendah jika dibandingkan dengan model tanpa SMOTE yang cenderung menyebar pada rentang 0,7510 hingga 0,8550.



Gambar 11 Perbandingan (a) akurasi (b) sensitivitas (c) spesifisitas (d) AUC

Setelah diterapkan metode SMOTE pada data latih untuk membangun model regresi logistik, nilai sensitivitas dari model klasifikasi menjadi meningkat. Meningkatnya nilai sensitivitas ini menjadi suatu prioritas untuk menentukan model

yang terbaik dalam mengklasifikasikan mahasiswa BUD IPB yang tidak lulus karena penelitian ini lebih memfokuskan model dalam mengklasifikasikan mahasiswa BUD IPB yang tidak lulus. Kesalahan mahasiswa BUD IPB yang tidak lulus diklasifikasikan sebagai mahasiswa BUD IPB yang lulus akan berakibat fatal karena tindakan pencegahan yang dilakukan agar mahasiswa BUD IPB tidak mengalami putus kuliah berisiko untuk tidak tepat sasaran. Hal ini dapat mengakibatkan angka putus kuliah mahasiswa BUD IPB dapat meningkat. Model klasifikasi regresi logistik biner dengan SMOTE ditentukan sebagai model yang terbaik dalam mengklasifikasikan mahasiswa BUD IPB yang tidak lulus karena mampu meningkatkan sensitivitas yang berarti mampu menurunkan kesalahan dalam mengklasifikasikan mahasiswa BUD IPB yang tidak lulus sebagai mahasiswa BUD IPB yang lulus.

#### 4.5 Peubah-peubah yang Berpengaruh terhadap Ketidaklulusan Mahasiswa BUD IPB

Model yang digunakan dalam mengklasifikasikan mahasiswa BUD IPB yang tidak lulus adalah model yang dibangun dari seluruh data yang telah melalui tahap SMOTE. Model dibangun dari tujuh peubah penjelas sebagai berikut:

$$g(x) = 3,277 - 0,579 X_{1(1)} - 3,842 X_{2(1)} - 1,099 X_{4(1)} + 1,356 X_{5(1)} \\ - 1,156 X_{6(1)} - 3,128 X_{6(2)} - 1,925 X_{6(3)} - 3,559 X_{6(4)} \\ + 0,056 X_7 - 2,631 X_8$$

Hasil pengujian parameter secara simultan dengan uji G menunjukkan bahwa nilai statistik uji G sebesar 1026 lebih besar dari  $\chi^2_{(0,05,7)} = 14,06$ , maka tolak  $H_0$  artinya terdapat peubah penjelas yang berpengaruh terhadap ketidaklulusan mahasiswa BUD IPB. Kemudian dilakukan uji *Wald* untuk melihat peubah mana saja yang berpengaruh terhadap ketidaklulusan mahasiswa BUD IPB. Hasil uji *Wald* pada Tabel 8, menunjukkan bahwa peubah penjelas yang berpengaruh terhadap ketidaklulusan mahasiswa BUD IPB pada taraf nyata 5% adalah  $X_1, X_2, X_4, X_5, X_6, X_7, X_8$ .

Tabel 8 Hasil uji Wald regresi logistik biner

Peubah	Koefisien	Galat Baku	<i>Wald</i>	<i>p-value</i>
Konstanta	3,277	1,153	8,082	0,004*
$X_{1(1)}$	-0,579	0,138	17,584	0,000*
$X_{2(1)}$	-3,842	0,428	80,661	0,000*
$X_{4(1)}$	-1,099	0,221	24,720	0,000*
$X_{5(1)}$	1,356	0,251	29,227	0,000*
$X_{6(1)}$	-1,156	0,165	48,845	0,000*
$X_{6(2)}$	-3,128	0,484	41,696	0,000*
$X_{6(3)}$	-1,925	1,089	3,125	0,077
$X_{6(4)}$	-3,559	1,114	10,210	0,001*
$X_7$	0,056	0,015	13,970	0,000*
$X_8$	-2,631	0,151	304,712	0,000*

Keterangan : \*) signifikan pada taraf nyata 5%

Rasio odds digunakan untuk menginterpretasikan koefisien dalam regresi logistik. Rasio odds merupakan rasio atau perbandingan dari peluang kejadian sukses terjadi dan kejadian sukses tidak terjadi. Kejadian sukses pada penelitian ini adalah mahasiswa BUD IPB yang tidak lulus. Interpretasi rasio odds dilakukan pada peubah penjelas yang berpengaruh signifikan terhadap model. Rasio odds yang digunakan untuk menginterpretasikan model disajikan pada Tabel 9.

Tabel 9 Nilai rasio odds

Peubah	Rasio Odds
$X_{1(1)}$	0,561
$X_{2(1)}$	0,021
$X_{4(1)}$	0,333
$X_{5(1)}$	3,881
$X_{6(1)}$	0,315
$X_{6(2)}$	0,044
$X_{6(4)}$	0,028
$X_7$	1,058
$X_8$	0,072

Peubah jenis kelamin memiliki rasio odds sebesar 0,561. Nilai ini menunjukkan mahasiswa BUD IPB berjenis kelamin perempuan memiliki kecenderungan tidak lulus 0,561 kali dibandingkan mahasiswa BUD IPB berjenis kelamin laki-laki. Interpretasi lainnya yaitu mahasiswa BUD IPB berjenis kelamin laki-laki memiliki kecenderungan lebih besar untuk tidak lulus dibandingkan mahasiswa BUD IPB berjenis kelamin perempuan. Hasil eksplorasi menunjukkan kejadian tidak lulus lebih banyak dialami mahasiswa BUD IPB laki-laki.

Peubah status sekolah asal mahasiswa BUD IPB memiliki rasio odds sebesar 0,021. Nilai ini menunjukkan mahasiswa BUD IPB yang berasal dari sekolah swasta memiliki kecenderungan tidak lulus sebesar 0,021 kali dibandingkan mahasiswa BUD IPB yang berasal dari sekolah negeri. Interpretasi lainnya yaitu mahasiswa BUD IPB yang berasal dari sekolah negeri memiliki kecenderungan yang lebih besar untuk tidak lulus dibandingkan mahasiswa BUD IPB yang berasal dari sekolah swasta. Hasil eksplorasi menunjukkan mahasiswa BUD IPB yang berasal dari sekolah negeri lebih banyak dibandingkan mahasiswa BUD IPB dari sekolah swasta dan kejadian tidak lulus lebih banyak dialami mahasiswa BUD IPB yang berasal dari sekolah negeri.

Peubah kelompok program studi memiliki rasio odds sebesar 0,333. Nilai ini menunjukkan mahasiswa BUD IPB dengan kelompok program studi sains dan sosial memiliki kecenderungan tidak lulus sebesar 0,333 kali dibandingkan mahasiswa BUD IPB dari kelompok program studi sains dan teknologi. Interpretasi lainnya yaitu mahasiswa BUD IPB dengan kelompok program studi sains dan teknologi memiliki kecenderungan yang lebih besar untuk tidak lulus dibandingkan mahasiswa BUD IPB dengan kelompok program studi sains dan sosial. Hasil eksplorasi menunjukkan mahasiswa BUD IPB dengan kelompok program studi sains dan teknologi lebih banyak dibandingkan mahasiswa BUD IPB dengan

kelompok program studi sains dan sosial serta kejadian tidak lulus lebih banyak dialami mahasiswa BUD IPB dengan kelompok program studi sains dan teknologi.

Peubah ada atau tidaknya PPU pada mahasiswa BUD IPB memiliki rasio odds sebesar 3,881. Nilai ini menunjukkan mahasiswa BUD IPB dengan PPU memiliki kecenderungan tidak lulus 3,881 kali dibandingkan mahasiswa BUD IPB tanpa PPU. Hasil eksplorasi menunjukkan kejadian tidak lulus lebih banyak dialami oleh mahasiswa BUD IPB dengan PPU.

Peubah tipe sponsor pemerintah kabupaten/kota memiliki rasio odds sebesar 0,315. Peubah tipe sponsor pemerintah provinsi memiliki rasio odds sebesar 0,044 sementara peubah tipe sponsor yayasan atau lembaga lainnya memiliki rasio odds sebesar 0,028. Nilai-nilai rasio odds tersebut menunjukkan mahasiswa BUD IPB dengan tipe sponsor pemerintah kabupaten/kota memiliki kecenderungan tidak lulus sebesar 0,315 kali dibandingkan mahasiswa BUD IPB dengan tipe sponsor perusahaan, mahasiswa BUD IPB dengan tipe sponsor pemerintah provinsi memiliki kecenderungan tidak lulus sebesar 0,044 dibandingkan mahasiswa BUD IPB dengan tipe sponsor perusahaan sementara mahasiswa BUD IPB dengan tipe sponsor yayasan atau lembaga lainnya memiliki kecenderungan tidak lulus sebesar 0,028 dibandingkan mahasiswa BUD IPB dengan tipe sponsor perusahaan. Interpretasi lainnya yaitu mahasiswa BUD IPB dengan tipe sponsor perusahaan memiliki kecenderungan yang lebih besar untuk tidak lulus dibandingkan mahasiswa BUD IPB dengan tipe sponsor pemerintah kabupaten/kota, pemerintah provinsi, dan yayasan atau lembaga lainnya. Hasil eksplorasi menunjukkan kejadian tidak lulus lebih banyak dialami oleh mahasiswa BUD IPB dengan tipe sponsor pemerintah kabupaten/kota dan perusahaan.

Peubah rata-rapor memiliki rasio odds sebesar 1,058. Nilai ini menunjukkan semakin besar rata-rapor yang dimiliki mahasiswa BUD IPB, maka kecenderungan tidak lulus mahasiswa BUD IPB lebih besar 1,058 kali. Hasil eksplorasi menunjukkan kejadian tidak lulus dialami mahasiswa BUD IPB dengan rata-rapor yang cenderung menyebar pada rentang 78,42 hingga 86,83 sementara kejadian lulus dialami mahasiswa BUD IPB dengan rata-rapor yang cenderung menyebar pada rentang 80,67 hingga 87,94. Selain itu, terdapat beberapa mahasiswa BUD IPB dengan status lulus memiliki rata-rapor dibawah 70.

Peubah IPK TPB/PPKU memiliki rasio odds sebesar 0,072. Nilai ini menunjukkan semakin besar IPK TPB/PPKU mahasiswa BUD IPB, maka kecenderungan mahasiswa BUD IPB untuk tidak lulus sebesar 0,072 kali. Interpretasi lainnya yaitu mahasiswa BUD IPB dengan IPK TPB/PPKU yang semakin kecil memiliki kecenderungan lebih besar untuk tidak lulus. Hasil eksplorasi menunjukkan kejadian tidak lulus dialami mahasiswa BUD IPB dengan IPK TPB/PPKU yang cenderung menyebar pada rentang 1,74 hingga 2,36 sementara kejadian lulus dialami mahasiswa BUD IPB dengan IPK TPB/PPKU yang cenderung menyebar pada rentang 2,46 hingga 3,29.



## **V SIMPULAN**

Model klasifikasi yang dibangun dari data latih dengan menerapkan metode SMOTE menghasilkan sensitivitas yang lebih tinggi dibandingkan model tanpa metode SMOTE. Meningkatnya nilai sensitivitas menandakan bahwa model mampu mengklasifikasikan mahasiswa BUD IPB yang tidak lulus (kelas minoritas) dengan lebih baik sehingga kesalahan dalam pengklasifikasian mahasiswa BUD IPB yang tidak lulus sebagai mahasiswa BUD IPB yang lulus menjadi rendah. Sementara untuk kriteria akurasi, spesifisitas, dan nilai AUC nilainya menurun. Penurunan ini terjadi karena adanya peningkatan mahasiswa BUD IPB yang lulus diklasifikasikan sebagai mahasiswa BUD IPB yang tidak lulus. Hal ini menunjukkan bahwa model sudah tidak cenderung memprediksi mahasiswa BUD IPB yang lulus (kelas mayoritas). Peubah-peubah yang berpengaruh terhadap ketidakiulusan mahasiswa BUD IPB yaitu jenis kelamin, status sekolah asal, kelompok program studi, ada atau tidaknya PPU, tipe sponsor, rata-rata rapor, dan IPK TPB/PPKU.

## DAFTAR PUSTAKA

- Admisi IPB. 2021. [diunduh pada Desember 2021]. <https://admisi.ipb.ac.id/>
- Agresti A. 2007. *Categorical Data Analysis*. Ed ke-2. New Jersey: John Wiley and Sons.
- Amelia OD, Soleh AM, Rahardiantoro S. 2018. Pemodelan *support vector machine* data tidak seimbang keberhasilan studi mahasiswa magister IPB. *Xplore*. 2(1):33-40.
- [BUD IPB] Beasiswa Utusan Daerah IPB. 2021. [diunduh pada Desember 2021]. <https://BUD IPB.ipb.ac.id/>
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. 2002. SMOTE (*Synthetic Minority Over-sampling Technique*). *Journal of Artificial Intelligence Research*. 16:321-357.
- Fox J, Monette G. 1992. Generalized collinearity diagnostics. *Journal of the American Statistical Association*. 87(417):178-183.
- Garson GD. 2008. *Logistik Regression*. [diunduh pada Desember 2021].
- Hair JF, Black WC, Babin BJ, Anderson RE. 2010. *Multivariate Data Analysis*. Ed ke-7. New York(US): Prentice Hall International.
- Hosmer DW, Lemeshow S, Sturdivant RX. 2013. *Applied Logistic Regression*. Ed ke-3. New York(US): John Wiley and Sons.
- Imran F, Susetyo B, Wigena AH. 2013. Identifikasi faktor-faktor yang berhubungan dengan mahasiswa putus kuliah di IPB angkatan 2008 menggunakan analisis *survival*. *Xplore*. 2(1):1-6.
- Jennrich RI, Robinson SM. 1969. A newton-raphson algorithm for maximum likelihood factor analysis. *Psychometrika*. 34(1):111-123.
- Musu W, Ibrahim A, Heriadi. 2021. Pengaruh komposisi data *training* dan *testing* terhadap akurasi algoritma C4.5. *Sistem Informasi dan Teknologi Informasi*. 10(1):186-195.
- Nurgenita N. 2015. Identifikasi faktor-faktor yang memengaruhi ketepatan waktu kelulusan mahasiswa program sarjana IPB [skripsi]. Bogor: Institut Pertanian Bogor.
- Osborne JW. 2007. *Best Practices in Quantitative Methods*. USA : SAGE Publications.
- Pangestika MP, Sumertajaya IM, Rizki A. 2021. Penerapan *Synthetic Minority Oversampling Technique* (SMOTE) pada pemodelan regresi logistik biner terhadap keberhasilan studi mahasiswa program magister IPB. *Xplore*. 10(2):152-166.
- Pramana S, Yuniarto B, Mariyah S, Santoso I, Nooraeni R. 2018. *Data Mining dengan R Konsep serta Implementasi*. Bogor: In Media.
- Provost F, Fawcett T. 2001. Robust classification for imprecise environments. *Machine Learning*. 42:203–231.
- Rahmah H. 2013. Penerapan SMOTE pada metode *cruise* untuk penentuan faktor keberhasilan studi mahasiswa BUD IPB [skripsi]. Bogor: Institut Pertanian Bogor.

- Senaviratna NAMR, Cooray TMJA. 2019. Diagnosing multicollinearity of logistic regression model. *Asian Journal of Probability and Statistics*. 5(2):1-9.
- Widasari E. 2017. Penerapan SMOTE pada pemodelan regresi logistik biner (studi kasus : data pekerja tahun 2014 di Kabupaten Badung) [skripsi]. Bogor: Institut Pertanian Bogor.

## RIWAYAT HIDUP

Penulis dilahirkan di kota Metro pada tanggal 16 Juli 2000 sebagai anak kedua dari pasangan bapak Aria Purnama dan ibu Eko Sari Kurniasih. Pendidikan sekolah menengah atas (SMA) ditempuh di sekolah SMA Al Kautsar Bandar Lampung dan lulus pada tahun 2018. Pada tahun 2018, penulis diterima sebagai mahasiswa program sarjana (S-1) di Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam IPB *University*.

Selama mengikuti program S-1, penulis aktif mengikuti kegiatan organisasi seperti Himpunan Profesi Gamma Sigma Beta (GSB) periode 2020/2021 sebagai anggota badan pengawas. Penulis juga aktif di beberapa kegiatan kepanitiaan seperti Kompetisi Statistika Junior 2020 sebagai bendahara divisi kompetisi, Statistika Ria dan Festival Sains Data (Satria Data) 2020 sebagai panitia yang membantu jalannya perlombaan *Statistics Essay Competition* (SEC), Kompetisi Statistika Junior 2021 sebagai anggota divisi kompetisi. Penulis juga pernah melaksanakan Praktik Lapang di Badan Pusat Statistik (BPS) Provinsi Lampung pada bulan Oktober-November 2021.