

Identifikasi Peubah yang Berpengaruh terhadap Ketidاكلulusan Mahasiswa Program Sarjana BUD IPB dengan Regresi Logistik Biner*

Mahdiyah Riaesnianda¹, Aam Alamudi^{1‡}, Agus M Soleh¹, and Septian Rahardiantoro¹

¹Department of Statistics, IPB University, Indonesia

[‡]corresponding author: aamalamudi@gmail.com

Copyright © 2023 Mahdiyah Riaesnianda, Aam Alamudi, Agus M Soleh, and Septian Rahardiantoro. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

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

* Received: Sep 2022; Reviewed: Dec 2022; Published: Jan 2023

1. Pendahuluan

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.

Peubah-peubah yang berpengaruh terhadap ketidakkulusan mahasiswa BUD IPB dapat dianalisis dengan regresi logistik biner. 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 mahasiswa BUD IPB. Oleh karena itu, penelitian ini akan menggunakan analisis regresi logistik biner untuk mengetahui peubah yang berpengaruh terhadap ketidakkulusan mahasiswa BUD IPB. Namun, 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 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%.

2. Metodologi

2.1 Bahan dan Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Direktorat Administrasi Pendidikan dan Penerimaan Mahasiswa Baru (DAPPMB) 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. Daftar peubah yang digunakan disajikan pada Tabel 1.

Tabel 1 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	-

2.2 Prosedur Analisis Data

Penelitian ini menggunakan perangkat lunak R 4.0.2 dengan tahapan-tahapan yang dilakukan sebagai berikut:

- 1) Melakukan pra-proses data yaitu melakukan seleksi terhadap mahasiswa BUD IPB yang memiliki data tidak lengkap, memetakan peubah akreditasi yang merujuk pada Badan Akreditasi Nasional Sekolah/Madrasah (BAN-S/M), mengelompokkan peubah jenis sponsor merujuk pada *website* BUD IPB, mengelompokkan peubah program studi menjadi dua, yaitu kelompok program studi Sains dan Teknologi (ST) serta kelompok Sains dan Sosial (SS).
- 2) Melakukan eksplorasi data untuk mengetahui gambaran umum mahasiswa BUD IPB berdasarkan status kelulusannya. Selanjutnya, melakukan pemeriksaan multikolinearitas pada peubah penjelas 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).

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

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.

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

- 3) Membagi data secara acak menjadi 80% data latih dan 20% data uji.
- 4) Melakukan analisis regresi logistik biner dengan tahapan sebagai berikut.
 - a. Membangun model dengan data latih. Menurut Hosmer *et al.* (2013), apabila terdapat peubah penjelas berskala nominal memiliki k kategori maka diperlukan $k-1$ peubah boneka (*dummy*) dalam model. Persamaan model:

$$g(\mathbf{x}) = \beta_0 + \sum_{i=1}^p \beta_i x_i + \sum_{l=1}^{k_i-1} \beta_{il} D_{il}$$

dengan β_0 merupakan konstanta, β_i merupakan koefisien regresi logistik, x_i merupakan peubah penjelas dengan $i = 1, \dots, p$, p merupakan banyaknya peubah penjelas, β_{il} merupakan koefisien peubah boneka pada peubah penjelas ke- i dan kategori ke- l , D_{il} merupakan peubah boneka pada peubah penjelas ke- i dan kategori ke- l , dan k_i merupakan banyaknya kategori pada peubah penjelas ke- i dengan $l = 1, \dots, k_i - 1$.

- b. Melakukan evaluasi model dengan *confusion matrix* dan nilai AUC dari data uji.
- 5) Melakukan SMOTE dengan tahapan sebagai berikut:
 - a. 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*.
 - b. Menghitung jarak antar amatan pada kelas minoritas menggunakan persamaan jarak *Euclidean*. Persamaan jarak *Euclidean*:

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

dengan $d(\mathbf{a}, \mathbf{b})$ merupakan jarak *Euclidean* untuk dua vektor amatan, a_i merupakan nilai amatan \mathbf{a} untuk peubah penjelas ke- i , b_i merupakan nilai amatan \mathbf{b} untuk peubah penjelas ke- i , p merupakan banyaknya peubah penjelas dengan $i = 1, 2, \dots, p$.

- c. Menentukan nilai k yaitu tiga untuk k -tetangga terdekat.
 - d. Satu contoh amatan dipilih secara acak dari kelas minoritas.

- e. Menentukan amatan k-tetangga terdekat dengan mengurut jarak contoh amatan terpilih dengan semua amatan pada kelas minoritas.
- f. Membangkitkan data buatan berdasarkan amatan k tetangga terdekat. Data numerik dibangkitkan dengan persamaan:
- $$x_{baru} = x + (x^* - x) \times rand [0,1]$$
- dengan x_{baru} adalah vektor amatan baru, x adalah vektor amatan awal, 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.
- g. Langkah d sampai f dilakukan hingga jumlah kelas minoritas dan kelas mayoritas menjadi seimbang.
- 6) 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.
- 7) Mengulangi tahap tiga sampai enam sebanyak seratus kali.
- 8) Membandingkan hasil evaluasi kedua model klasifikasi (tanpa tahap SMOTE dan melalui tahap SMOTE) dengan melihat *boxplot* dari *confusion matrix* dan nilai AUC.
- 9) Membangun model dengan data keseluruhan berdasarkan hasil perbandingan evaluasi model pada tahap 8:
- a. Menduga parameter regresi logistik biner dengan pendekatan metode kemungkinan maksimum. Pendugaan parameter β dilakukan dengan memaksimumkan persamaan *log-likelihood* sebagai berikut:
- $$L(\beta) = \ln [l(\beta)] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln [1 - \pi(x_i)]\}$$
- Nilai $\hat{\beta}$ diperoleh dengan menurunkan $L(\beta)$ terhadap $\beta_i = 0$ secara iteratif dengan metode Newton-Raphson (Jennrich dan Robinson 1969).
- b. Melakukan pengujian signifikansi parameter secara simultan atau keseluruhan menggunakan statistik uji G:
- $$G = -2 \ln \frac{L_0}{L_p} \sim \chi^2_{(ab)}$$
- dengan L_0 adalah fungsi kemungkinan maksimum tanpa peubah penjelas dan L_p adalah fungsi kemungkinan maksimum dengan peubah penjelas.
- c. Melakukan pengujian signifikansi parameter secara parsial menggunakan statistik uji wald:
- $$W_i = \left[\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \right]^2 \sim \chi^2_{(ab)}$$
- dengan $\hat{\beta}_i$ adalah nilai dugaan untuk parameter β_i , $SE(\hat{\beta}_i)$ adalah galat baku untuk $\hat{\beta}_i$
- 10) Interpretasi model berdasarkan nilai rasio *odds* dari peubah-peubah yang berpengaruh signifikan. Menurut Hosmer et al. (2013), rasio odds dilambangkan dengan OR. Hubungan rasio odds dengan koefisien regresi adalah sebagai berikut.

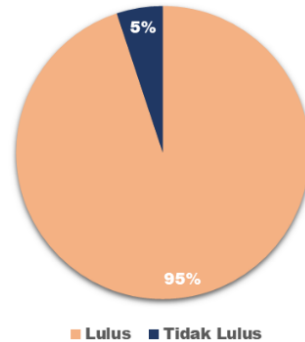
$$OR = \exp(\beta_i)$$

3. Hasil dan Pembahasan

3.1 Eksplorasi Data

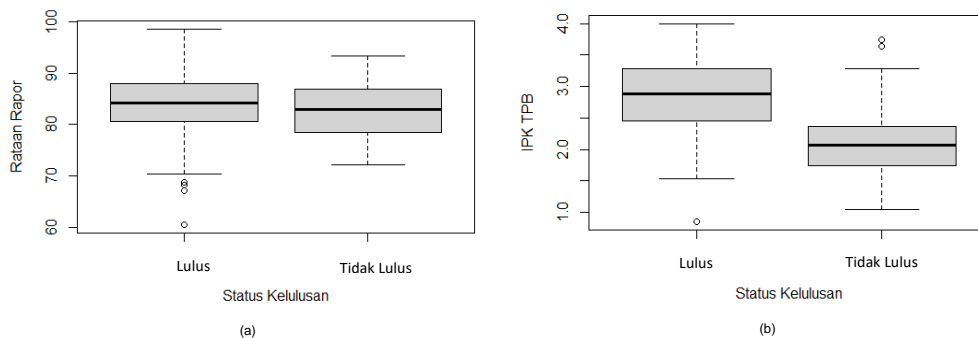
Eksplorasi data dilakukan untuk melihat gambaran umum mahasiswa BUD IPB berdasarkan status kelulusannya. Data yang digunakan sebanyak 925 mahasiswa BUD IPB yang sudah memiliki status kelulusan. Persentase status kelulusan mahasiswa BUD IPB dapat dilihat pada Gambar 1.

Gambar 1 menunjukkan mahasiswa BUD IPB yang berstatus lulus lebih banyak dibandingkan mahasiswa BUD IPB yang berstatus tidak lulus. Perbedaan persentase yang jauh mengindikasikan adanya ketidakseimbangan data antara kelas lulus dan tidak lulus.



Gambar 1 Persentase status kelulusan mahasiswa BUD IPB

Peubah penjelas numerik yang digunakan adalah rata-ran rapor dan IPK TPB/PPKU. Gambar 2(a) menunjukkan mayoritas mahasiswa BUD IPB yang lulus memiliki rata-ran rapor 80,68 hingga 87,94. Rataan rapor maksimum mahasiswa BUD IPB yang berstatus lulus yaitu 98,60 dan mahasiswa BUD IPB yang berstatus tidak lulus yaitu 93,33.

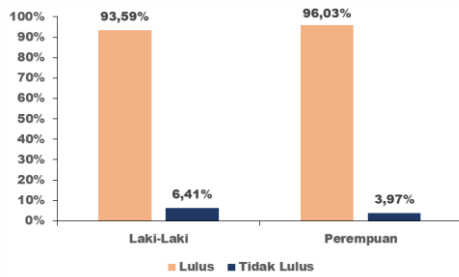


Gambar 2 *Boxplot* peubah penjelas (a) rata-ran rapor (b) IPK TPB/PPKU

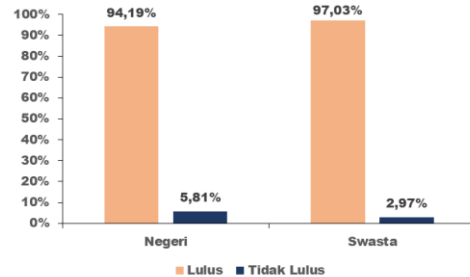
Gambar 2(b) menunjukkan median IPK TPB/PPKU mahasiswa BUD IPB yang berstatus lulus lebih besar dibandingkan mahasiswa BUD IPB yang tidak lulus. Mahasiswa BUD IPB yang berstatus lulus memiliki nilai IPK TPB/PPKU yang lebih beragam dibandingkan mahasiswa BUD IPB yang tidak lulus.

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. 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. Persentase status kelulusan mahasiswa BUD IPB berdasarkan akreditasi sekolah asal terlihat pada Gambar 5.

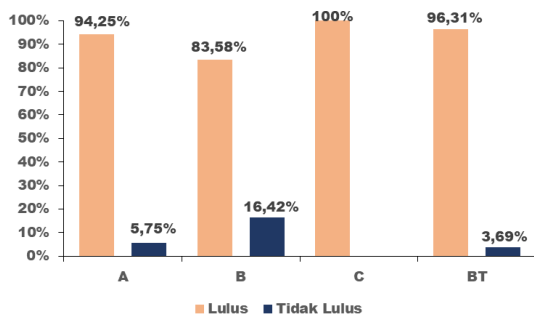
Mahasiswa BUD IPB yang berasal dari sekolah Belum Terakreditasi (BT) persentasenya lebih banyak dibandingkan sekolah yang berasal dari akreditasi lainnya, yaitu sebesar 67,35%.



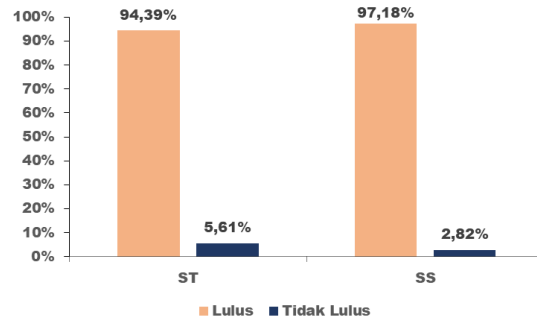
Gambar 3 Persentase status kelulusan berdasarkan jenis kelamin



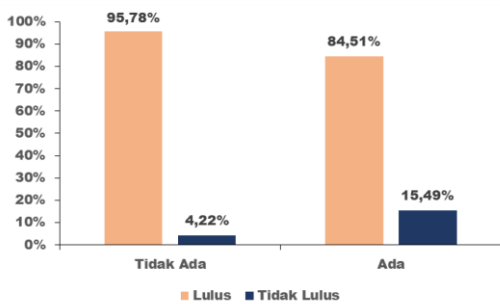
Gambar 4 Persentase status kelulusan berdasarkan status sekolah asal



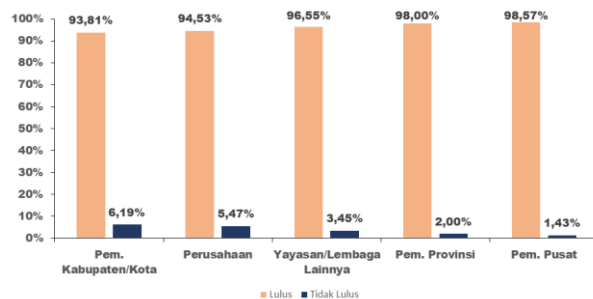
Gambar 5 Persentase status kelulusan berdasarkan akreditasi sekolah



Gambar 6 Persentase status kelulusan berdasarkan kelompok prodi



Gambar 7 Persentase status kelulusan berdasarkan PPU



Gambar 8 Persentase status kelulusan berdasarkan tipe sponsor

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%. Persentase status kelulusan berdasarkan PPU BUD IPB ditunjukkan pada Gambar 7. Mahasiswa BUD IPB yang tidak mengikuti PPU lebih banyak sebesar 92,32%. 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.

Langkah pertama sebelum dilakukan analisis regresi logistik biner yaitu pemeriksaan apakah terdapat masalah multikolinearitas yang serius antar peubah

penjelas atau tidak. Seperti menurut Fox dan Monette (1992), $GVIF \frac{1}{2 \times db}$ yang dikuadratkan digunakan untuk memeriksa multikolinearitas pada peubah penjelas. Tabel 2 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 2 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

3.2 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 3.

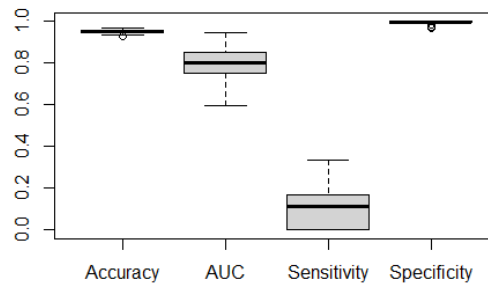
Tabel 3 Banyaknya data sebelum SMOTE

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 disajikan pada Gambar 9.

Gambar 9 menunjukkan akurasi, spesifisitas dan AUC yang dihasilkan tinggi sementara sensitivitas yang dihasilkan model 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 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

3.3 Regresi Logistik Biner dengan SMOTE

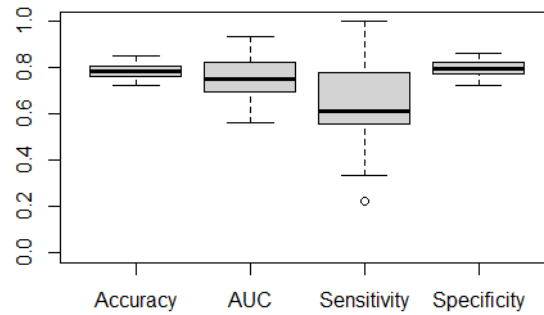
SMOTE dilakukan dengan membangkitkan data baru dengan menerapkan *k-nearest neighbours* untuk setiap data di kelas minoritas. Pembangkitan data dilakukan hingga banyaknya amatan mahasiswa BUD IPB pada kelas yang lulus sama dengan mahasiswa BUD IPB yang tidak lulus. Persentase banyaknya amatan mahasiswa 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 4.

Tabel 4 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.

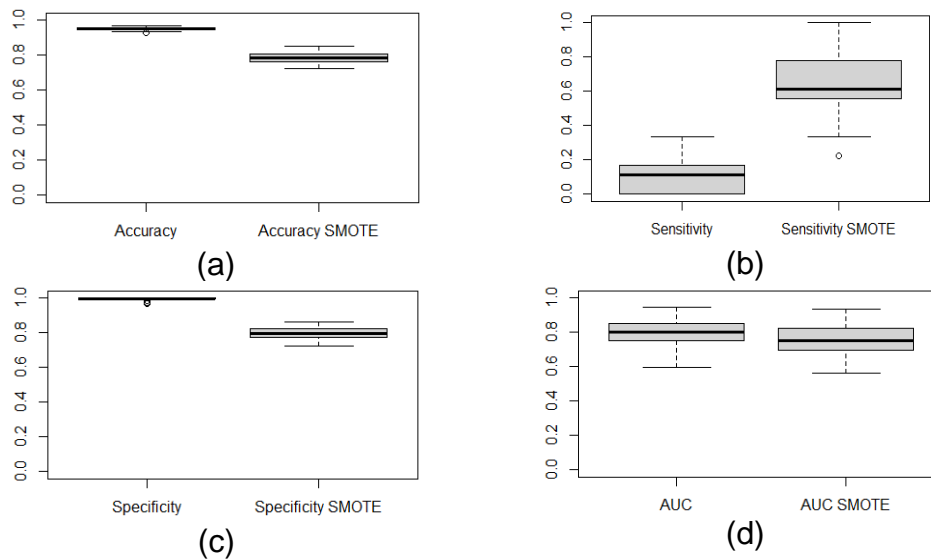
Gambar 10 menunjukkan sensitivitas yang dihasilkan model dengan SMOTE meningkat dibandingkan model tanpa SMOTE. Hal tersebut memperkuat bahwa model dengan SMOTE mampu mengklasifikasikan mahasiswa BUD IPB yang tidak lulus dengan baik karena kesalahan mahasiswa BUD IPB yang tidak lulus diklasifikasikan sebagai mahasiswa BUD IPB yang lulus rendah. 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 sensitivitas yang meningkat dan menurunnya akurasi, spesifisitas, dan nilai AUC.



Gambar 10 Kinerja klasifikasi model regresi logistik dengan SMOTE

3.4 Perbandingan Kinerja Klasifikasi Regresi Logistik Biner

Perbandingan akurasi, sensitivitas, spesifisitas, dan AUC dari model regresi logistik biner tanpa SMOTE dan dengan SMOTE ditunjukkan pada Gambar 11. 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.



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

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.

3.5 Peubah-peubah yang Berpengaruh terhadap Ketidاكلulusan 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 ketidاكلulusan mahasiswa BUD IPB. Kemudian dilakukan uji *Wald* untuk melihat peubah mana saja yang berpengaruh terhadap ketidاكلulusan mahasiswa BUD IPB. Hasil uji *Wald* pada Tabel 5, menunjukkan bahwa peubah penjelas yang berpengaruh terhadap ketidاكلulusan mahasiswa BUD IPB pada taraf nyata 5% adalah $X_1, X_2, X_4, X_5, X_6, X_7, X_8$.

Tabel 5 Hasil uji Wald regresi logistik biner

Peubah	Koefisien	Galat Baku	Wald	p-value
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 6.

Tabel 6 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. 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. 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. 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.

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. Hasil eksplorasi menunjukkan kejadian tidak lulus lebih banyak dialami oleh mahasiswa BUD IPB dengan tipe sponsor pemerintah kabupaten/kota dan perusahaan.

Peubah rata-rata rapor memiliki rasio odds sebesar 1,058. Nilai ini menunjukkan semakin besar rata-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-rata rapor yang cenderung menyebar pada rentang 78,42 hingga 86,83 sementara kejadian lulus dialami mahasiswa BUD IPB dengan rata-rata rapor yang cenderung menyebar pada rentang 80,67 hingga 87,94.

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. 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.

4. 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 menunjukkan 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 ketidakkulusan mahasiswa BUD IPB yaitu jenis kelamin, status sekolah asal, kelompok program studi, ada atau tidaknya PPU, tipe sponsor, rataan rapor, dan IPK TPB/PPKU.

Daftar Pustaka

- Admisi IPB. 2021. [diunduh pada Desember 2021]. <https://admisi.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.
- 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.
- 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.
- 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.
- Widasari E. 2017. Penerapan SMOTE pada pemodelan regresi logistik biner (studi kasus : data pekerja tahun 2014 di Kabupaten Badung) [skripsi]. Bogor: Institut Pertanian Bogor.