

PENGEMBANGAN MODEL PREDIKSI KELULUSAN CALON MAHASISWA SARJANA PADA SISTEM SELEKSI SNMPTN IPB

WADUDI MUTHAHARI



**PROGRAM MAGISTER ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Perpustakaan IPB University



PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Pengembangan Model Prediksi Kelulusan Calon Mahasiswa Sarjana pada Sistem Seleksi SNMPTN IPB” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apapun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

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

Bogor, Juni 2025

Wadudi Muthahari
G6501211053



RINGKASAN

WADUDI MUTHAHARI. Pengembangan Model Prediksi Kelulusan Calon Mahasiswa Sarjana pada Sistem Seleksi SNMPTN IPB. Dibimbing oleh SONY HARTONO WIJAYA dan UTAMI DYAH SYAFITRI.

Institut Pertanian Bogor (IPB) merupakan institusi pendidikan yang berperan penting dalam pembangunan sumberdaya manusia, salah satunya dengan melakukan seleksi penerimaan mahasiswa baru melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN). SNMPTN merupakan pola seleksi berdasarkan potensi akademik pelamar selama di sekolah pada masing-masing Perguruan Tinggi Negeri (PTN) di bawah koordinasi Lembaga Tes Masuk Perguruan Tinggi (LTMPT). Sejak tahun 2019, proses seleksi di IPB menggunakan media seleksi berbasis web dan algoritma tertentu. Pemodelan berbasis *machine learning* dapat diterapkan untuk mempercepat proses seleksi dengan memberikan rekomendasi apakah siswa yang mendaftar layak diterima atau tidak sebagai mahasiswa IPB. Setiap tahun data pelamar yang lulus dan tidak lulus selalu memiliki ketidakseimbangan karena pelamar yang lulus jauh lebih sedikit dibandingkan dengan jumlah pelamar yang tidak lulus. Teknik penyeimbangan data yang dilakukan adalah *Synthetic Minority Oversampling Technique* (SMOTE). Penelitian ini bertujuan mengembangkan pemodelan *machine learning* menggunakan *Random Forest* (RF) dan Regresi Logistik Biner (RLB) pada sistem seleksi SNMPTN IPB dan mencari faktor-faktor apa saja yang mempengaruhi kelulusan calon mahasiswa jalur SNMPTN IPB.

Penelitian ini menggunakan data SNMPTN 2021 sebagai data latih dan SNMPTN 2022 sebagai data uji. Sebanyak empat skenario pemodelan diterapkan pada masing-masing metode dan wilayah. Semua model dilatih menggunakan *hyperparameter tuning*. Model kesatu menggunakan semua fitur dan tanpa penyeimbangan. Model kedua menggunakan semua fitur dan SMOTE. Model ketiga menggunakan seleksi fitur dan SMOTE serta model keempat dengan pemilihan fitur by Adjustment Pakar (AP) dan SMOTE. Fokus utama dalam pemodelan dan pengujian adalah memaksimalkan deteksi pelamar yang lulus, sehingga nilai *recall* yang tinggi menjadi acuan dalam pengambilan model terbaik. Hasil pengujian menunjukkan bahwa model yang diuji menggunakan data uji dengan penyeimbangan data SMOTE secara konsisten menunjukkan nilai *recall* yang lebih tinggi dibandingkan dengan model tanpa penyeimbangan data. Model ketiga dengan Regresi Logistik Biner pada data Jawa Barat dan model kedua dengan Regresi Logistik Biner pada data Non Jawa Barat menunjukkan nilai *recall* terbaik sebesar 88,93% dan 86,91%. Hasil pemodelan juga menunjukkan bahwa urutan pemilihan perguruan tinggi, indeks sekolah, pilihan program studi, dan prestasi memiliki pengaruh signifikan terhadap prediksi kelulusan pelamar.

Kata kunci: Model, *Random Forest*, Regresi Logistik Biner, SMOTE, SNMPTN IPB.

SUMMARY

WADUDI MUTHAHARI. Development of a Graduation Prediction Model for Undergraduate Applicants in the SNMPTN Admission System at IPB University. Supervised by SONY HARTONO WIJAYA and UTAMI DYAH SYAFITRI.

IPB University is an educational institution that plays a vital role in human resource development, one of which is by conducting new student admission selection through the SNMPTN pathway. SNMPTN is a selection pattern based on the academic potential of applicants while at school at each State University (PTN) under the coordination of the Higher Education Entrance Test Institute (LTMPT). Since 2019, the selection process at IPB has used web-based selection media and specific algorithms. Machine learning-based modeling can be utilized to streamline the selection process by providing recommendations on the eligibility of applicants for admission as students at IPB. Every year, the data on applicants who pass and fail always shows an imbalance, as the number of applicants who pass is significantly smaller than the number of applicants who fail. The data balancing technique used is the Synthetic Minority Oversampling Technique (SMOTE). This study aims to develop machine learning models using Random Forest (RF) and Binary Logistic Regression (BLR) in the IPB SNMPTN selection system and identify the factors that influence the graduation of prospective students through the IPB SNMPTN pathway.

The research uses SNMPTN 2021 data as training data and SNMPTN 2022 as test data. There are four models built. All models are trained using hyperparameter tuning. The first model uses all features without balancing. The second model uses all features and SMOTE. The third model utilizes feature selection and SMOTE, while the fourth model employs feature selection via Expert Adjustment (EA) and SMOTE. The primary focus in modeling and testing is to maximize the detection of successful applicants, so that high recall values are used as a reference in selecting the best model. The test results show that models tested using test data with SMOTE data balancing consistently yield higher recall values compared to models without data balancing. The third model, with Binary Logistic Regression applied to West Java data, and the second model, with Binary Logistic Regression applied to Non-West Java data, show the best recall values of 88,93% and 86,91%, respectively. The modeling results also show that the order of college selection, school index category, choice of study program, and academic achievement have a significant influence on the prediction of applicants' passing rates.

Keywords: *Binary Logistic Regression*, IPB SNMPTN, Model, *Random Forest*, SMOTE.



@Hak cipta milik IPB University

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta Milik IPB, tahun 2025
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.

PENGEMBANGAN MODEL PREDIKSI KELULUSAN CALON MAHASISWA SARJANA PADA SISTEM SELEKSI SNMPTN IPB

WADUDI MUTHAHARI

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister pada
Program Magister Ilmu Komputer

**PROGRAM MAGISTER ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

IPB University

Penguji Luar Komisi Pembimbing pada Ujian Tesis:
Dr. Toto Haryanto, S.Kom., M.Si.



IPB University
— Bogor Indonesia —

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Judul Tesis : Pengembangan Model Prediksi Kelulusan Calon Mahasiswa
Sarjana pada Sistem Seleksi SNMPTN IPB
Nama : Wadudi Muthahari
NIM : G6501211053

Disetujui Oleh

Pembimbing 1:

Dr. Sony Hartono Wijaya, S.Kom., M.Kom.



Pembimbing 2:

Dr. Utami Dyah Syafitri, M.Si.



Diketahui Oleh

Ketua Program Studi:

Dr. Toto Haryanto, S.Kom., M.Si.
NIP. 19821117 201404 1 001



Dekan Sekolah Sains Data, Matematika, dan Informatika:

Prof. Dr. Ir. Agus Buono, M.Si., M.Kom.
NIP. 19660702 199302 1 001



Tanggal Ujian: 24 Maret 2025

Tanggal Lulus:



PRAKATA

Puji serta syukur penulis panjatkan kepada Allah SWT atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan karya ilmiah dengan judul “Pengembangan Model Prediksi Kelulusan Calon Mahasiswa Sarjana pada Sistem Seleksi SNMPTN IPB”. Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister Ilmu Komputer, Institut Pertanian Bogor. Penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

Dosen Pembimbing Dr. Sony Hartono Wijaya, S.Kom., M.Kom. dan Dr. Utami Dyah Syafitri, M.Si. yang telah dengan sabar dan ketelatenan membimbing dan mengarahkan penulis dalam penyusunan karya ilmiah ini.

Penguji luar komisi pembimbing, perwakilan program studi, pembahas kolokium, moderator kolokium, dan moderator seminar.

- Ibu, ayah, istri, anak dan seluruh keluarga yang selalu mendukung dan menyemangati penulis.
- Teman-teman seangkatan di Departemen Ilmu Komputer yang selalu saling memberikan dukungan
- Semua pihak yang telah membantu penulis dalam menyelesaikan karya ilmiah ini.
- Penulis menyadari bahwa karya ilmiah ini masih jauh dari sempurna. Oleh karena itu, penulis menerima kritik dan saran yang membangun demi perbaikan karya ilmiah ini dimasa depan.

Semoga karya ilmiah ini dapat bermanfaat bagi Kementerian Pertanian dalam meningkatkan tata kelola keamanan informasi dalam menunjang layanan informasi pertanian.

Bogor, Juni 2025

Wadudi Muthahari

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR ISI

DAFTAR TABEL	ix
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	ix
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup	4
II TINJAUAN PUSTAKA	5
2.1 Regresi Logistik Biner	5
2.2 <i>Random Forest</i>	6
2.3 <i>Boruta Feature Selection</i> (BFS)	7
2.4 <i>Synthetic Minority Oversampling Technique</i> (SMOTE)	8
2.5 <i>K-Fold Cross Validation</i>	9
2.6 Evaluasi Model	10
2.7 Sistem Seleksi SNMPTN	11
III METODE	12
3.1 Data Penelitian	12
3.2 Praproses Data	14
3.3 Seleksi Fitur menggunakan BFS	14
3.4 Penyeimbangan Data menggunakan SMOTE	16
3.5 Pemodelan	16
3.6 Pengujian dan Analisis Hasil Pengujian	17
3.7 Hasil Pengujian pada Sistem Seleksi SNMPTN	17
3.8 Lingkungan Pengembangan	17
IV HASIL DAN PEMBAHASAN	18
4.1 Data Penelitian	18
4.2 Praproses Data	19
4.3 Seleksi Fitur menggunakan BFS	20
4.4 SMOTE	21
4.5 Pemodelan	21
4.6 Pengujian dan Analisis Hasil Pengujian	23
4.7 Analisis Faktor-Faktor yang Mempengaruhi Kelulusan	30
4.8 Hasil Pengujian pada Sistem Seleksi SNMPTN	31
V SIMPULAN DAN SARAN	32
5.1 Simpulan	32
5.2 Saran	32
DAFTAR PUSTAKA	33
LAMPIRAN	36
RIWAYAT HIDUP	49



DAFTAR TABEL

1 <i>Confusion Matrix</i>	10
2 Fitur-fitur pada <i>dataset</i> seleksi SNMPTN	13
3 Metode praproses data pada fitur-fitur seleksi SNMPTN	14
4 Jumlah percobaan model menggunakan <i>hyperparameter tuning</i>	16
5 Distribusi jumlah data latih (2021) dan data uji (2022)	18
6 Hasil praproses <i>imputation</i> dan <i>encoding</i>	19
7 Distribusi jumlah data latih (2021) setelah penyeimbangan	21
8 Hasil pemodelan data Jawa Barat dan Non Jawa Barat	22
9 Hasil pengujian performa model data Jawa Barat dan Non Jawa Barat	23
10 Prediksi model terhadap pelamar yang lulus aktual	29
11 Fitur-fitur yang mempengaruhi kelulusan	30

DAFTAR GAMBAR

1 Proses algoritma <i>Random Forest</i> (Wu <i>et al.</i> 2019)	7
2 Ilustrasi cara kerja algoritma BFS (Hasan <i>et al.</i> 2020)	8
3 Ilustrasi SMOTE	9
4 Tampilan beranda sistem seleksi SNMPTN	11
5 Tampilan seleksi sistem seleksi SNMPTN	11
6 Tahapan penelitian	12
7 Persentase lulus dan tidak lulus data latih dan data uji	18
8 Fitur penting berdasarkan nilai rata-rata <i>Z-score</i> tertinggi	20
9 Grafik perbandingan akurasi model Jawa Barat dan Non Jawa Barat	24
10 Grafik perbandingan <i>precision</i> dan <i>recall</i> data Jawa Barat	24
11 Grafik perbandingan <i>precision</i> dan <i>recall</i> data Non Jawa Barat	25
12 Histogram distribusi peluang data Jawa Barat	26
13 Histogram distribusi peluang data Non Jawa Barat	28
14 Tampilan media seleksi SNMPTN dengan menampilkan nilai peluang	31

DAFTAR LAMPIRAN

1 Hasil model satu dan dua RF data Jawa Barat	37
2 Hasil model tiga dan empat RF data Jawa Barat	38
3 Hasil model satu dan dua RF data Non Jawa Barat	39
4 Hasil model tiga dan empat RF data Non Jawa Barat	40
5 Hasil model satu dan dua RLB data Jawa Barat	41
6 Hasil model tiga dan empat RLB data Jawa Barat	43
7 Hasil model satu dan dua RLB data Non Jawa Barat	45
8 Hasil model tiga dan empat RLB data Non Jawa Barat	47