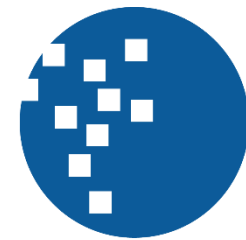


**PERBANDINGAN MODEL MACHINE LEARNING UNTUK
PREDIKSI STATUS KELULUSAN AKADEMIK MAHASISWA
DENGAN INTERPRETASI SHAP**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

LAPORAN RESEARCH & TECHNOLOGY PROGRAM

King Mahaputra Batara

00000083179

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2026

**PERBANDINGAN MODEL MACHINE LEARNING UNTUK
PREDIKSI STATUS KELULUSAN AKADEMIK MAHASISWA
DENGAN INTERPRETASI SHAP**



LAPORAN RESEARCH & TECHNOLOGY PROGRAM

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer (S.Kom)

King Mahaputra Batara

00000083179

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2026

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : King Mahaputra Batara
Nomor Induk Mahasiswa : **00000083179**
Program Studi : Sistem Informasi

Laporan Research and Techonolgy Program dengan judul:

PERBANDINGAN MODEL MACHINE LEARNING UNTUK PREDIKSI
STATUS KELULUSAN AKADEMIK MAHASISWA DENGAN
INTERPRETASI SHAP

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 19 Desember 2025



King Mahaputra Batara

HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : King Mahaputra Batara
NIM : 00000083179
Program Studi : Sistem Informasi
Judul Laporan : Perbandingan Model *Machine Learning* untuk
Prediksi Status Kelulusan Akademik Mahasiswa dengan Interpretasi
SHAP

Dengan ini saya menyatakan secara jujur menggunakan bantuan Kecerdasan Artifisial (AI) dalam pengerjaan ~~Tugas/Laporan/Project/Tugas Akhir~~ sebagai berikut:

- ☒ Menggunakan AI sebagaimana diizinkan untuk membantu dalam menghasilkan ide-ide utama saja
- ☒ Menggunakan AI sebagaimana diizinkan untuk membantu menghasilkan teks pertama saja
- ☒ Menggunakan AI untuk menyempurnakan sintaksis dan tata bahasa untuk pengumpulan tugas
- ☐ Karena tidak diizinkan: Tidak menggunakan bantuan AI dengan cara apa pun dalam pembuatan tugas

Saya juga menyatakan bahwa:

- (1) Menyerahkan secara lengkap dan jujur penggunaan perangkat AI yang diperlukan dalam tugas melalui Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI)
- (2) Mengakui telah menggunakan bantuan AI dalam tugas saya baik dalam bentuk kata, paraphrase, penyertaan ide atau fakta penting yang disarankan oleh AI dan saya telah menyantumkan dalam sitasi serta referensi
- (3) Terlepas dari pernyataan di atas, tugas ini sepenuhnya merupakan karya saya sendiri

Tangerang, 19 Desember 2025



King Mahaputra Batara

HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN

Yang bertanda tangan di bawah ini:

Nama : King Mahaputra Batara

NIM : 00000083179

Program Studi : Sistem Informasi

Fakultas : Teknik dan Informatika

menyatakan bahwa saya melaksanakan kegiatan di:

Nama Perusahaan/Organisasi : Universitas Multimedia Nusantara

Alamat : Jalan Scientia Boulevard Gading, Curug Sangereng,
Serpong, Kabupaten Tangerang, Banten 15810

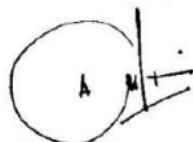
Email Perusahaan/Organisasi : <https://www.umn.ac.id/>

1. Perusahaan/Organisasi tempat saya melakukan kegiatan dapat di validasi keberadaannya.
2. Jika dikemudian hari, terbukti ditemukan kecurangan/penyimpangan data yang tidak valid di perusahaan/organisasi tempat saya melakukan kegiatan, maka:
 - a. Saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.
 - b. Saya bersedia menerima semua sanksi yang berlaku sebagaimana ditetapkan dalam peraturan yang berlaku di Universitas Multimedia Nusantara.

Pernyataan ini saya buat dengan sebenar-benarnya dan digunakan sebagaimana mestinya.

Tangerang, 19 Desember 2025

Mengetahui



Assoc Prof. Dr. Ir.
Andrianingsih, S.Kom.,MMSI

Menyatakan



King Mahaputra Batara

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : King Mahaputra Batara
NIM : 00000083179
Program Studi : Sistem Informasi
Jenjang : S1
Judul Karya Ilmiah : Perbandingan Model *Machine Learning*
untuk Prediksi Status Kelulusan Akademik Mahasiswa dengan Interpretasi
Shap

Menyatakan dengan sesungguhnya bahwa saya bersedia:

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 19 Desember .2025



King Mahaputra Batara

KATA PENGANTAR

Puji Syukur atas selesainya penulisan Laporan *Research and Technology Program* ini dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer (S.Kom) Jurusan Sistem Informasi Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan laporan penelitian PROSTEP ini, sangatlah sulit bagi saya untuk menyelesaikan laporan penelitian PROSTEP ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Ir. Andrey Andoko, M.Sc., Ph.D, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Suryasari, S.Kom., M.T., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya laporan penelitian PROSTEP ini.
5. Assoc Prof. Dr. Ir. Andrianingsih, S.Kom., MMSI, sebagai *supervisor* dari penelitian ini yang telah memberikan bimbingan dan arahan, dan motivasi hingga selesainya laporan ini.
6. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan laporan penelitian PROSTEP ini.

Semoga karya ilmiah ini bermanfaat sebagai moda informasi terhadap pengembangan sistem prediksi status kelulusan mahasiswa dengan bantuan interpretasi model.

Tangerang, 19 Desember 2025



King Mahaputra Batara

PERBANDINGAN MODEL MACHINE LEARNING UNTUK PREDIKSI STATUS KELULUSAN AKADEMIK MAHASISWA DENGAN INTERPRETASI SHAP

King Mahaputra Batara

ABSTRAK

Penelitian ini bertujuan untuk membangun dan mengevaluasi model prediksi status kelulusan mahasiswa sebagai dasar pengembangan sistem peringatan dini akademik di Program Studi Sistem Informasi Universitas Multimedia Nusantara. Permasalahan utama yang dihadapi adalah keterlambatan deteksi risiko akademik akibat pemantauan yang masih bersifat reaktif, sehingga intervensi sering dilakukan ketika kondisi mahasiswa sudah memburuk. Penelitian ini menggunakan data akademik historis mahasiswa angkatan 2020–2024 yang diperoleh dari Biro Informasi Akademik, dengan total populasi 1.024 mahasiswa dan sampel terlabel sebanyak 421 mahasiswa yang telah menyelesaikan studi. Prediksi dilakukan untuk empat kategori status kelulusan, yaitu Lulus Lebih Awal, Lulus Tepat Waktu, Tidak Lulus Tepat Waktu, dan *Dropout*. Tiga algoritma *machine learning* digunakan dan dibandingkan, yaitu CatBoost, *Logistic Regression* L1 (Lasso), dan *K-Nearest Neighbors* (KNN), dengan penanganan ketidakseimbangan kelas menggunakan SMOTE. Hasil evaluasi menunjukkan bahwa model CatBoost memberikan kinerja terbaik dengan akurasi sebesar 69,05 persen dan nilai *F1-score* makro sebesar 0,593, mengungguli Lasso dan KNN yang masing-masing memiliki *F1-score* makro di bawah 0,55. Model terbaik kemudian diinterpretasikan menggunakan metode SHAP untuk mengidentifikasi kontribusi fitur-fitur akademik terhadap prediksi, sehingga hasil model dapat dijelaskan secara transparan dan relevan untuk konteks bimbingan akademik. Temuan penelitian ini menunjukkan bahwa pendekatan berbasis *machine learning* yang disertai interpretabilitas mampu memberikan prediksi yang lebih informatif dan dapat dimanfaatkan sebagai dasar pengambilan keputusan preventif dalam pengelolaan akademik mahasiswa.

Kata kunci: *CatBoost, Early Warning System, Kelulusan Mahasiswa, K-Nearest Neighbor, Lasso, Machine Learning, SHAP*

COMPARISON OF THREE MODELS FOR PREDICTING STUDENT ACADEMIC GRADUATION STATUS USING ACADEMIC DATA WITH SHAP-BASED INTERPRETATION

King Mahaputra Batara

ABSTRACT (English)

This study aims to develop and evaluate a student graduation status prediction model as a foundation for an academic early warning system in the Information Systems Study Program at Universitas Multimedia Nusantara. The main problem addressed is the delayed detection of academic risk due to monitoring practices that remain largely reactive, causing interventions to be implemented only after students' academic conditions have significantly declined. This research utilizes historical academic records of students from the 2020–2024 cohorts obtained from the Academic Information Bureau, with a total population of 1,024 students and a labeled sample of 421 students who have completed their studies. The prediction task focuses on four graduation status categories: Early Graduation, On-Time Graduation, Late Graduation, and Dropout. Three machine learning algorithms are implemented and compared, namely CatBoost, L1 Logistic Regression (Lasso), and K-Nearest Neighbors (KNN), with class imbalance handled using the Synthetic Minority Oversampling Technique (SMOTE). The evaluation results indicate that CatBoost achieves the best performance, with an accuracy of 69.05 percent and a macro F1-score of 0.593, outperforming both Lasso and KNN, whose macro F1-scores remain below 0.55. The best-performing model is further interpreted using SHAP to identify the contribution of academic features to the prediction outcomes, enabling transparent and explainable results that are relevant for academic advising. Overall, the findings demonstrate that an interpretable machine learning approach can provide informative predictions and effectively support preventive, data-driven decision-making in academic management.

Keywords: *Academic Early Warning System, CatBoost, K-Nearest Neighbor, Lasso, Machine Learning, SHAP, Student Graduation Status*

DAFTAR ISI

HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
KATA PENGANTAR.....	vi
ABSTRAK	vii
<i>ABSTRACT (English)</i>	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR PERSAMAAN.....	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Urgensi Penelitian	5
1.5 Luaran Penelitian.....	7
1.6 Manfaat Penelitian	8
BAB II LANDASAN TEORI	11
2.1 Penelitian Terdahulu	11
2.2 Tinjauan Teori.....	20
2.2.1 Kinerja Akademik Mahasiswa.....	20
2.2.2 Fenomena Inflasi IPK	21
2.2.3 Faktor Internal dan Eksternal yang Mempengaruhi Prestasi Akademik.....	22
2.2.4 Confusion Matrix	23
2.3 <i>Framework</i> dan Algoritma	25
2.3.1 Cross-Industry Standard Process for Data Mining (CRISP-DM)	25
2.3.2 CatBoost.....	27
2.3.3 Logistic Regression L1 (Lasso)	29

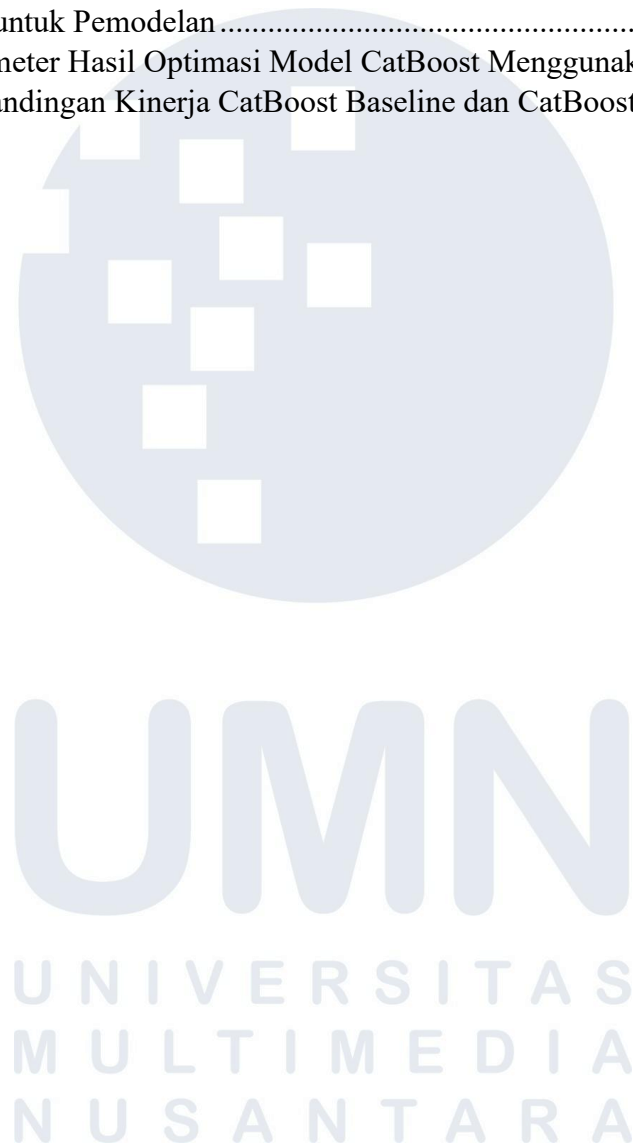
2.3.4	K-Nearest Neighbor	30
2.3.5	Hyperopt	31
2.4	Tools	32
2.4.1	Python	32
2.4.2	Google Collaboratory.....	33
BAB III METODOLOGI PENELITIAN		35
3.1	Gambaran Umum Objek Penelitian.....	35
3.2	Tahapan Penelitian	36
3.2.1	Business Understanding	36
3.2.2	Data Understanding	38
3.2.3	Data Preprocessing.....	40
3.2.4	Modeling	44
3.2.5	Evaluation	50
3.2.6	Deployment	52
3.3	Teknik Pengumpulan Data	54
3.3.1	Periode Pengambilan Data	54
3.4	Teknik Analisis Data.....	56
BAB IV ANALISIS DAN HASIL PENELITIAN		60
4.1	Hasil Penelitian.....	60
4.1.1	Business Understanding	60
2.2.1	Data Understanding	63
4.1.2	Data Preprocessing.....	73
4.1.3	Modeling	87
4.1.4	Evaluation	93
4.1.5	Deployment	104
4.2	Pembahasan Penelitian.....	105
4.2.1	Kinerja Model dalam Memprediksi Status Kelulusan	105
4.2.2	Perbandingan dan Hasil Evaluasi dari Ketiga Model Prediksi	108
4.2.3	Interpretasi Model Terbaik dengan SHAP	110
4.2.4	Implikasi dan Relevansi Penelitian	115
4.2.5	Keterbatasan dan Arah Penelitian Selanjutnya.....	116
BAB V SIMPULAN DAN SARAN		120

5.1	Simpulan	120
5.2	Saran	122
DAFTAR PUSTAKA		124
LAMPIRAN.....		128
A.	HASIL TURNITIN	128
B.	FORM KONSULTASI BIMBINGAN	138
C.	PRO-STEP 01	139
D.	PRO-STEP 02	140
E.	PRO-STEP 04	141



DAFTAR TABEL

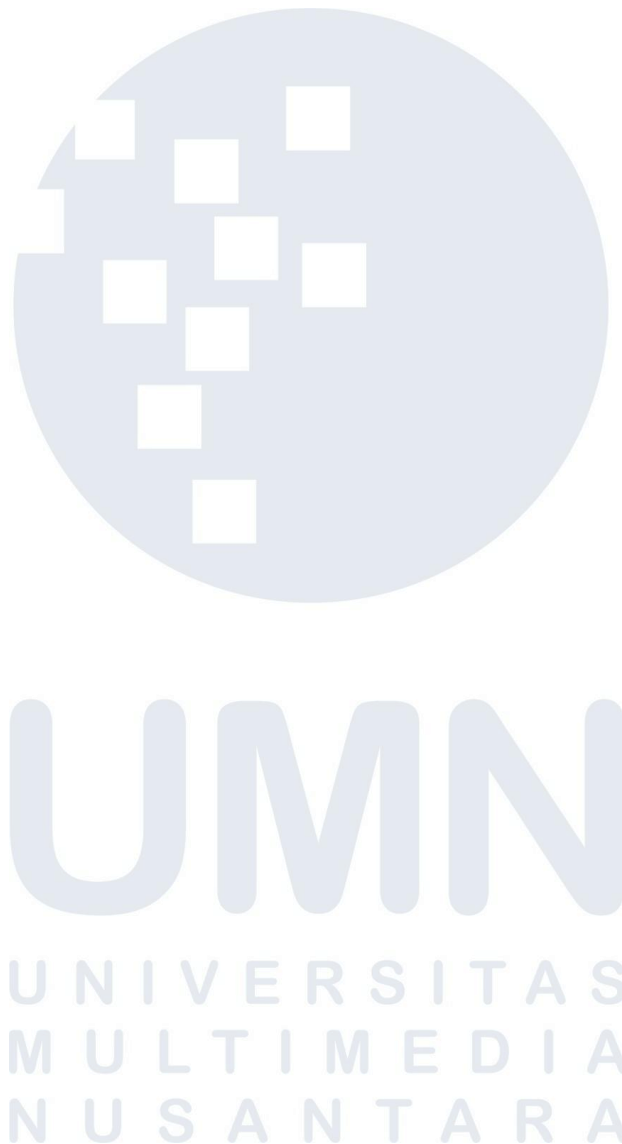
Tabel 21 Penelitian Terdahulu	12
Tabel 3.1 Hyperparameter Tiga Model Pembandingan	44
Tabel 3.2 Parameter Optimasi Model Catboost	49
Tabel 3.3 Parameter Tetap Model CatBoost	49
Tabel 3.4 Perbandingan Explainable AI	57
Tabel 4.1 Business Understanding	62
Tabel 4.2 Fitur untuk Pemodelan	64
Tabel 4. 3 Parameter Hasil Optimasi Model CatBoost Menggunakan Hyperopt .	92
Tabel 4. 4 Perbandingan Kinerja CatBoost Baseline dan CatBoost Teroptimasi	103



DAFTAR GAMBAR

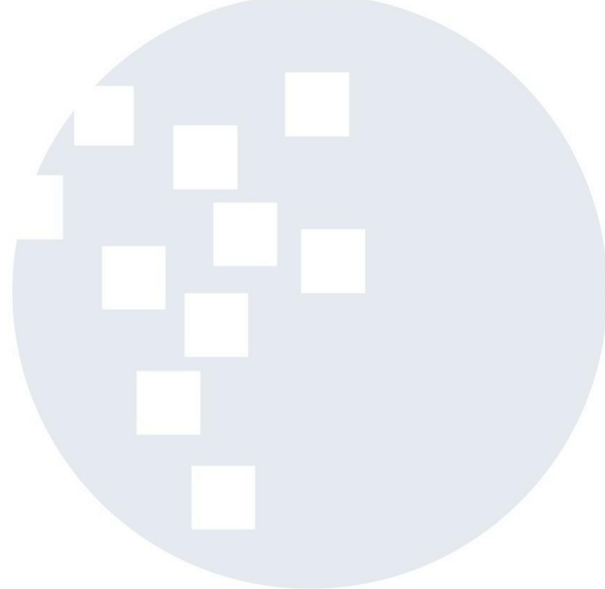
Gambar 2.1 Confusion Matric	24
Gambar 2.2 Alur CRISP-DM.....	26
Gambar 4.1 Distribusi Mahasiswa per Angkatan	65
Gambar 4.2 Rerata IPK per Angkatan	66
Gambar 4.3 Tren Rerata IPS per Semester	67
Gambar 4.4 Mata Kuliah dengan Mahasiswa Terbanyak	68
Gambar 4.5 Mata Kuliah dengan Rerata Nilai Rendah.....	68
Gambar 4.6 Rerata Nilai Mata Kuliah Wajib SI Terendah.....	69
Gambar 4.7 Rerata Nilai Mata Kuliah Wajib SI Tertinggi	70
Gambar 4. 8 Distribusi Status Kelulusan Mahasiswa	71
Gambar 4.9 Status Kelulusan Angkatan 2020	72
Gambar 4.10 Status Kelulusan Angkatan 2021	72
Gambar 4. 11 Status Kelulusan Angkatan 2022	73
Gambar 4. 12 Persentase Missing Values	74
Gambar 4.13 Contoh Data Nilai Angka Kosong	74
Gambar 4.14 Nilai NaN pada SKS, Nilai Angka, dan Nilai Huruf	75
Gambar 4.15 Data yang Sudah Bersih dari Missing Value	75
Gambar 4.16 Pseudocode Penentuan Status Kelulusan	76
Gambar 4.17 Data yang Berhasil Diidentifikasi Status Kelulusan	77
Gambar 4.18 Struktur Data Setelah Seleksi dan Transformasi Fitur	78
Gambar 4.19 Pipeline Preprocessing pada Tahap Pemodelan	79
Gambar 4.20 Distribusi Status Kelulusan	80
Gambar 4.21 Distribusi Slope IPS	81
Gambar 4.22 Distribusi Jumlah SKS Gagal.....	81
Gambar 4.23 Distribusi Jumlah Mata Kuliah Mengulang	82
Gambar 4.24 Distribusi Rerata Kehadiran Mahasiswa	83
Gambar 4.25 Distribusi IPS Terendah	83
Gambar 4.26 Distribusi Status Kelulusan pada Training.....	84
Gambar 4.27 Fitur Hasil Rekayasa	84
Gambar 4.28 Pembagian Data Sebelum Pemodelan.....	86
Gambar 4.29 Penerapan SMOTE pada Data Pelatihan.....	87
Gambar 4.30 Confusion Matrix Catboost.....	94
Gambar 4.31 Classification Report	95
Gambar 4.32 Confusion Matrix Lasso	96
Gambar 4.33 Classification Report Lasso.....	98
Gambar 4. 34 Confusion Matrix KNN.....	99
Gambar 4.35 Classification report KNN.....	100
Gambar 4. 36 Classification Report Model Selection.....	101
Gambar 4.37 Hasil Pemilihan Fitur Terbaik	101
Gambar 4.38 Ranking Fitur Terpenting oleh SHAP	110
Gambar 4.39 Faktor Penentu Kelas Dropout oleh SHAP	111

Gambar 4.40 Faktor Penentu Kelas Lulus Lebih Awal oleh SHAP	112
Gambar 4. 41 Faktor Penentu Kelas Lulus Tepat Waktu oleh SHAP	112
Gambar 4.42 Faktor Penentu Kelas Tidak Lulus Tepat Waktu oleh SHAP	113
Gambar 4.43 Contoh Force Plot SHAP untuk Mahasiswa Dropout.....	114



DAFTAR PERSAMAAN

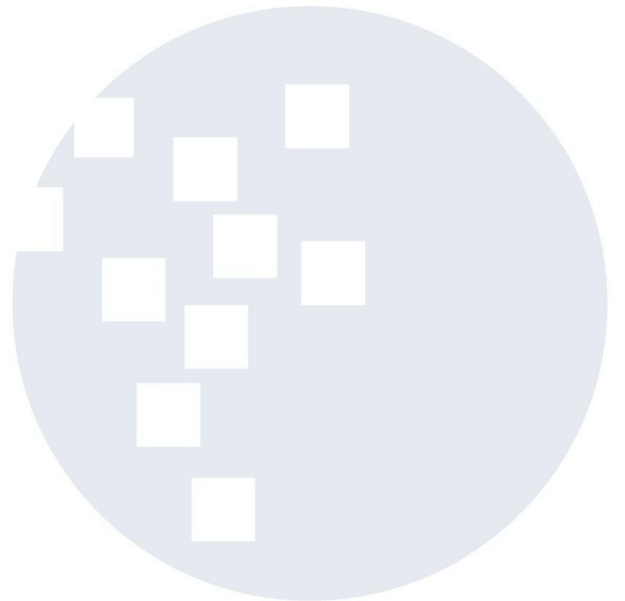
Rumus 2.1 Akurasi.....	24
Rumus 2.2 Presisi.....	24
Rumus 2.3 Recall	25
Rumus 2.4 F1 Score	25
Persamaan 2.5 CatBoost dengan iterasi ke-M.....	28
Persamaan 2.6 MSE untuk meminimalkan L.....	28



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran 1 Hasil Turnitin.....	137
Lampiran 2 Form Konsultasi Bimbingan.....	138
Lampiran 3 PRO-STEP 01	139
Lampiran 4 PRO-STEP 02.....	140
Lampiran 5 PRO-STEP 03.....	140
Lampiran 6 PRO-STEP 04.....	141



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA