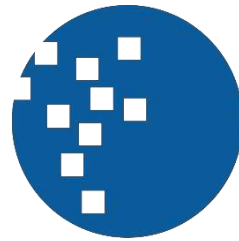


**OPTIMASI CATBOOST DENGAN INTERPRETASI SHAP
UNTUK PREDIKSI KINERJA AKADEMIK MAHASISWA
SISTEM INFORMASI UNIVERSITAS MULTIMEDIA
NUSANTARA**



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Laporan Skripsi

King Mahaputra Batara

00000083179

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

2025

**OPTIMASI CATBOOST DENGAN INTERPRETASI SHAP
UNTUK PREDIKSI KINERJA AKADEMIK MAHASISWA
SISTEM INFORMASI UNIVERSITAS MULTIMEDIA
NUSANTARA**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer

King Mahaputra Batara

00000083179

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

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : King Mahaputra Batara

Nomor Induk Mahasiswa **00000083179**

Program Studi : Sistem Informasi

Skripsi dengan judul:

**OPTIMASI CATBOOST DENGAN INTERPRETASI SHAP
UNTUK PREDIKSI KINERJA AKADEMIK MAHASISWA
SISTEM INFORMASI UNIVERSITAS MULTIMEDIA
NUSANTARA**

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, 28 November 2025



King Mahaputra Batara

LEMBAR 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 Skripsi : Optimasi Catboost dengan Interpretasi Shap untuk
Prediksi Kinerja Akademik Mahasiswa Sistem Informasi Universitas
Multimedia Nusantara

Dengan ini saya menyatakan secara jujur menggunakan bantuan Kecerdasan Artifisial (AI) dalam pengerjaan ~~Tugas/Project~~ Tugas Akhir*(coret salah satu) sebagai berikut (beri tanda centang yang sesuai):

- ☒ 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) Saya mengakui bahwa saya 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

Terlepas dari pernyataan di atas, tugas ini sepenuhnya merupakan karya saya sendiri

Tangerang, 28 November 2025
Yang 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 : Optimasi Catboost dengan Interpretasi Shap
untuk Prediksi Kinerja Akademik Mahasiswa Sistem Informasi Universitas
Multimedia Nusantara

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, 28 November 2025



King Mahaputra Batara

KATA PENGANTAR

Penulisan tugas akhir ini dapat diselesaikan berkat bimbingan, dukungan, dan bantuan dari berbagai pihak. Ucapan terima kasih disampaikan kepada dosen pembimbing, para pengajar, serta pihak program studi yang telah memberikan arahan selama proses penelitian berlangsung. Penghargaan juga diberikan kepada keluarga dan rekan-rekan yang selalu memberikan dukungan. Penelitian ini disusun untuk memenuhi syarat akademik serta diharapkan dapat memberikan manfaat bagi pengembangan sistem pendukung keputusan di lingkungan akademik. Semoga hasil yang diperoleh dapat menjadi referensi bagi penelitian selanjutnya dan berkontribusi pada peningkatan kualitas pemantauan kinerja mahasiswa.

Mengucapkan terima kasih

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 Universitas Multimedia Nusantara.
3. Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Universitas Multimedia Nusantara.
4. Suryasari, S.Kom., M.T., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga karya ilmiah ini dapat memberikan kontribusi bagi pengembangan sistem pendukung keputusan akademik, khususnya dalam upaya meningkatkan ketepatan intervensi dan pemantauan kinerja studi mahasiswa.

Tangerang, 28 November 2025



King Mahaputra Batara

OPTIMASI CATBOOST DENGAN INTERPRETASI SHAP UNTUK PREDIKSI KINERJA AKADEMIK MAHASISWA SISTEM INFORMASI UNIVERSITAS MULTIMEDIA NUSANTARA

King Mahaputra Batara

ABSTRAK

Penelitian ini bertujuan membangun sistem peringatan dini untuk memprediksi status kelulusan mahasiswa Program Studi Sistem Informasi UMN dengan memanfaatkan algoritma CatBoost pada tiga titik evaluasi akademik yaitu Semester 2, 4, dan 6. Tantangan utama yang dihadapi adalah ukuran data yang terbatas serta ketidakseimbangan kelas yang cukup ekstrem, terutama pada kategori *dropout* dan tidak lulus tepat waktu. Untuk mengatasi hal ini, penelitian menerapkan serangkaian teknik *balancing* seperti SMOTE, SMOTE+Tomek, serta optimasi *hyperparameter* menggunakan Optuna dengan skema *cross validation*. Hasil menunjukkan bahwa performa model meningkat secara signifikan setelah *balancing* dan *tuning* dengan model terbaik mencapai akurasi hingga 0.76 dan F1-Macro 0.71 pada Semester 6. Analisis SHAP digunakan untuk mengidentifikasi fitur yang paling berpengaruh dan memberikan interpretasi terhadap keputusan model, di mana faktor akademik seperti IPS, IPK, jumlah SKS gagal, serta pola kehadiran terbukti konsisten menentukan prediksi. Temuan ini menunjukkan bahwa pendekatan *machine learning* yang tepat dapat memberikan dukungan analitis yang kuat bagi institusi pendidikan dalam mendeteksi risiko mahasiswa secara lebih dini dan terukur.

Kata kunci: CatBoost, *Early Warning System*, Ketidakseimbangan Kelas, *Machine Learning*, Prediksi Akademi

OPTIMIZING CATBOOST WITH SHAP INTERPRETATION FOR PREDICTION ACADEMIC PERFORMANCE OF INFORMATION SYSTEMS STUDENTS AT UNIVERSITAS MULTIMEDIA NUSANTARA

King Mahaputra Batara

ABSTRACT (English)

This study develops an Early Warning System (EWS) to predict the graduation outcomes of Information Systems students at Universitas Multimedia Nusantara using CatBoost as the main classification model. The research evaluates model performance across three academic checkpoints (Semester 2, 4, and 6) and addresses challenges related to small sample size and extreme class imbalance through resampling techniques such as SMOTE and SMOTE+Tomek, followed by extensive hyperparameter tuning using Optuna. The results show that model performance improves as students progress academically, with the Semester 6 model achieving the highest accuracy of 0.76 and an F1-Macro of 0.71. SHAP analysis provides insight into the most influential academic and non-academic features, highlighting the role of attendance patterns, GPA trends, failed credits, and family support indicators in shaping final predictions. Overall, the findings demonstrate that CatBoost can support early and late-stage academic risk detection and offer practical guidance for academic intervention strategies.

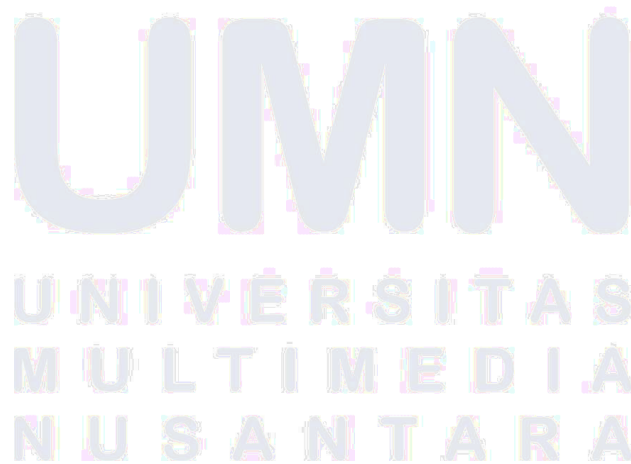
Keywords: CatBoost, Early Warning System, Machine Learning, SHAP, Student Performance Prediction

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR	v
ABSTRAK.....	vi
<i>ABSTRACT (English)</i>	vii
DAFTAR ISI	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR RUMUS.....	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	6
1.3 Batasan Masalah.....	7
1.4 Tujuan dan Manfaat Penelitian	9
1.4.1 Tujuan Penelitian	9
1.4.2 Manfaat Penelitian	9
1.5 Sistematika Penulisan	12
BAB II LANDASAN TEORI.....	13
2.1 Penelitian Terdahulu.....	13
2.2 Teori yang berkaitan.....	21
2.2.1 Faktor-Faktor Kinerja Akademik	21
2.2.2 Motivasi Belajar dan Keterlibatan Mahasiswa	22
2.2.3 Sistem Peringatan Dini (<i>Early Warning Systems</i> - EWS) dalam Pendidikan	22
2.3 <i>Framework</i> dan Algoritma.....	23
2.3.1 Educational Data Mining (EDM).....	24
2.3.2 <i>Cross-Industry Standard Process for Data Mining</i> (CRISP-DM)	24
2.3.3 Klasifikasi dalam Supervised Learning.....	26
2.3.4 Model <i>Gradient Boosting</i> dan <i>Catboost</i>	29

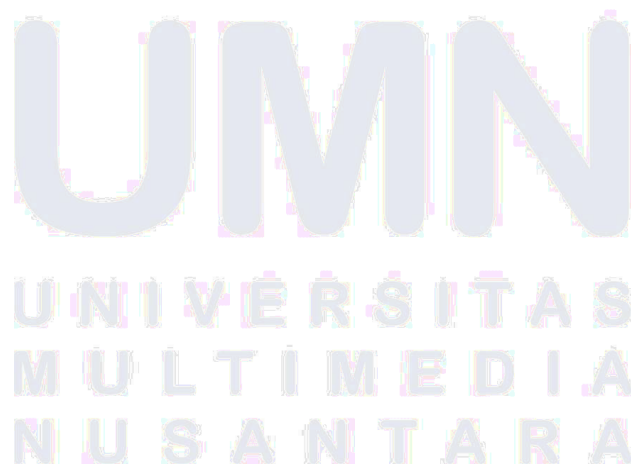
2.3.5	Penanganan Data Tidak Seimbang.....	33
2.3.6	Optimasi Model (Hyperparameter Tuning dan Optuna).....	39
2.3.7	Interpretasi Model: Explainable AI (XAI) dan SHAP.....	43
2.3.8	Evaluation Metrics	47
2.4	Tools.....	49
2.4.1	Python.....	50
2.4.2	Google Collaboratory	50
BAB III METODOLOGI PENELITIAN.....		52
3.1	Gambaran Umum Objek Penelitian	52
3.2	Tahapan Penelitian.....	53
3.2.1	Business Understanding.....	54
3.2.2	Data Understanding	56
3.2.3	Data Preprocessing	59
3.2.4	Modeling.....	66
3.2.5	Evaluation	81
3.2.6	Deployment	84
3.3	Teknik Pengumpulan Data.....	85
3.3.1	Periode Pengambilan Data	86
3.3.2	Populasi	86
3.3.3	Sampel	87
3.4	Teknik Analisis Data	89
BAB IV HASIL DAN PEMBAHASAN PENELITIAN		93
4.1	Hasil Penelitian	93
4.1.1	Business Understanding.....	93
4.1.2	Data Understanding	96
4.1.3	Data Preprocessing	113
4.1.4	Modeling.....	130
4.1.5	Evaluation	140
4.1.6	Deployment	160
4.2	Pembahasan Hasil Penelitian	163
4.2.1	Performa Model Catboost dalam Memprediksi Status Kelulusan Mahasiswa pada Data <i>Imbalance</i>	164

4.2.2 Dampak Optimasi Catboost dalam Mendukung EWS di Prodi Sistem Informasi UMN.....	165
4.2.3 Pola Prediksi dan Akurasi Sistem Peringatan Dini (EWS)	168
4.2.4 Faktor Akademik Paling Berpengaruh dengan SHAP terhadap Hasil Prediksi Catboost	171
4.2.5 Implikasi dan Relevansi.....	178
4.2.6 Keterbatasan dan Penelitian Lanjutan.....	179
BAB V SIMPULAN DAN SARAN	184
5.1 Simpulan.....	184
5.2 Saran	185
DAFTAR PUSTAKA	187
LAMPIRAN	198
Lampiran A Turnitin Similarity Report	198
Lampiran B Form Konsultasi Bimbingan.....	209



DAFTAR TABEL

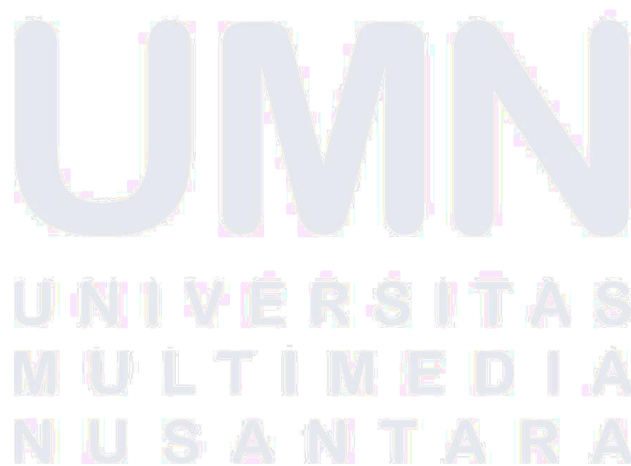
Tabel 2.1 Penelitian Terdahulu.....	14
Tabel 3.1 Hyperparameter Tuning Model Semester 4	70
Tabel 3. 2 Hyperparameter Tuning Model Semester 6	77
Tabel 3.3 Hyperparameter Tuning pada Pemodelan Sems. 6 dengan Data Eksternal.....	80
Tabel 3.4 Perbandingan Bahasa Pemrograman	90
Tabel 4.1 Business Understanding EWS.....	93
Tabel 4.2 Fitur dalam Pemodelan.....	97
Tabel 4.3 Karakteristik Kelas Status Kelulusan Mahasiswa	107
Tabel 4.4 Ketersediaan Data Akademik per Snapshot Semester.....	108
Tabel 4.5 Hyperparameter Tuning dengan Optuna Model Sems. 2.....	131
Tabel 4.6 Hasil Hyperparameter Terbaik pada Setiap Versi Pemodelan Sems. 4	135
Tabel 4.7 Hasil Hyperparameter Tuning pada Pemodelan Sems. 6.....	138
Tabel 4.8 Classification Metrics Semua Versi Pemodelan Sems. 4.....	147
Tabel 4.9 Hasil Optimasi Pemodelan Terbaik (Versi 2) Sems. 4.....	148
Tabel 4.10 Classification Metrics Semua Versi Pemodelan Sems. 6.....	154
Tabel 4.11 Hasil Optimasi Pemodelan Terbaik (Versi 1) Sems. 6.....	155



DAFTAR GAMBAR

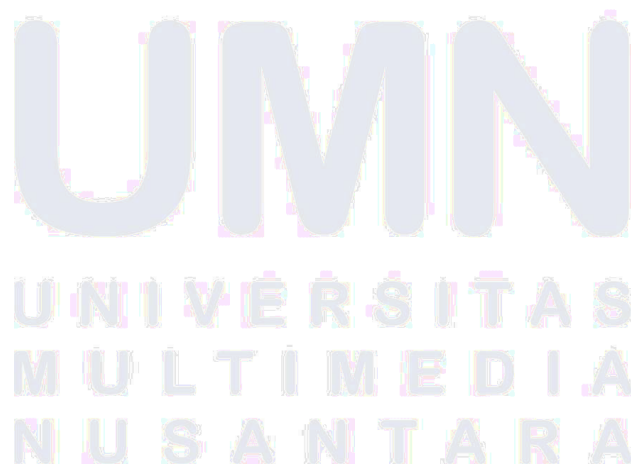
Gambar 2.1 Alur CRISP-DM [44]	25
Gambar 2.2 Confusion Matrix [114].....	48
Gambar 4.1 Distribusi Mahasiswa per Angkatan.....	101
Gambar 4.2 Rerata IPK per Angkatan.....	101
Gambar 4.3 Tren Rerata IPS per Semester.....	102
Gambar 4.4 Mata Kuliah dengan Mahasiswa Terbanyak.....	103
Gambar 4.5 Mata Kuliah dengan Rerata Nilai Rendah.....	103
Gambar 4. 6 Rerata Nilai Mata Kuliah Wajib SI Terendah	104
Gambar 4.7 Rerata Nilai Mata Kuliah Wajib SI Tertinggi.....	105
Gambar 4.8 Distribusi Status Kelulusan Mahasiswa.....	106
Gambar 4.9 Status Kelulusan Angkatan 2020.....	108
Gambar 4.10 Status Kelulusan Angkatan 2021.....	109
Gambar 4.11 Status Kelulusan Angkatan 2021.....	110
Gambar 4.12 Frekuensi Dukungan Keluarga	111
Gambar 4.13 Dukungan Finansial Keluarga	111
Gambar 4.14 Kepuasan Prodi Pilihan Sendiri.....	112
Gambar 4.15 Distribusi Kepuasa Prodi.....	113
Gambar 4.16 Persentase Missing Values	115
Gambar 4.17 Contoh Data Nilai Angka Kosong.....	115
Gambar 4.18 Nilai Nan pada SKS, Nilai Angka, dan Nilai Huruf.....	116
Gambar 4.19 Data yang sudah bersih dari Missing Value	116
Gambar 4.20 Psudo Code untuk Status Kelulusan	118
Gambar 4.21 Data yang Berhasil Diidentifikasi Status Kelulusan.....	119
Gambar 4.22 Hasil Encoding pada Data Eksternal	121
Gambar 4.23 Penerapan OHE pada Pemodelan Sems. 2	121
Gambar 4.24 Distribusi Status Kelulusan	122
Gambar 4. 25 Distribusi Slop IPS	123
Gambar 4.26 Distribusi Jumlah SKS Gagal	124
Gambar 4.27 Distribusi Jumlah Mata Kuliah Mengulang.....	124
Gambar 4.28 Distribusi Rerata Kehadiran Mahasiswa	125
Gambar 4.29 Distribusi IPS Terendah	125
Gambar 4.30 Distribusi Status Kelulusan pada Training	126
Gambar 4.31 Fitur Hasil Rekayasa.....	126
Gambar 4.32 Pembagian Data Sebelum Pemodelan	127
Gambar 4.33 Penerapan SMOTE pada Pemodelan Sems. 2.....	128
Gambar 4. 34 Penerapan SMOTE pada Pemodelan Sems. 4.....	129
Gambar 4. 35 Penerapan SMOTETomek pada Pemodelan Sems. 6.....	129
Gambar 4.36 Training vs Validation Loss Pemodelan Sems. 2.....	133
Gambar 4.37 Classification Report Pemodelan Sems. 2.....	142
Gambar 4.38 Confusion Metrics Pemodelan Sems. 2.....	143
Gambar 4.39 Confusion Metric Model Sems. 2 Kelas "Aman"	143

Gambar 4. 40 Confusion Metric Model Sems. 2 Kelas "Dropout"	144
Gambar 4.41 Hasil Evaluasi Model Terbaik Pemodelan Sems. 4.....	149
Gambar 4.42 Confusion Matrix Pemodelan Terbaik Sems. 4.....	150
Gambar 4.43 Penghentian Training Versi 2 Pemodelan Sems. 6	152
Gambar 4.44 Penghentian Training Versi 3 Pemodelan Sems. 6	152
Gambar 4.45 Hasil Evaluasi Model Terbaik Pemodelan Sems. 6.....	156
Gambar 4.46 Confusion Matrix Pemodelan Terbaik Sems. 6.....	157
Gambar 4.47 Hasil Evaluasi Model Terbaik Pemodelan Sems. 6 (Data Eksternal	159
Gambar 4.48 Confusion Matrix Pemodelan Sems. 6 dengan Data Eksternal	159
Gambar 4.49 Tampilan Awal EWS.....	161
Gambar 4.50 Hasil Prediksi Status Kelulusan Mahasiswa oleh EWS	162
Gambar 4.51 Interpretasi SHAP pada EWS.....	163
Gambar 4.52 Beeswarm SHAP pada Pemodelan CatBoost.....	172
Gambar 4.53 Force Plot SHAP pada Pemodelan CaBoost	173
Gambar 4.54 Force Plot SHAP TLTW Sems. 4.....	174
Gambar 4.55 Force Plot Dropout Pemodelan Sems. 4.....	174
Gambar 4.56 Force Plot SHAP Dropout Pemodelan Sems. 6.....	176
Gambar 4.57 Force Plot SHAP LLA Sems. 6.....	176
Gambar 4. 58 Force Plot SHAP LTW.....	177
Gambar 4.59 Force Plot SHAP TLTW Pemodelan Sems. 6.....	177



DAFTAR RUMUS

Rumus 2.1 Boosting [69].....	30
Rumus 2.2 Pembaruan Gradient [70]	30
Rumus 2.3 Ordered Boosting [72].....	31
Rumus 2.4 Encoding Catboost [73].....	32
Rumus 2.5 Log Loss [77]	32
Rumus 2.6 CatBoost.....	33
Rumus 2.7 SMOTE [79]	35
Rumus 2.8 <i>Tomek Link</i> [83]	36
Rumus 2.9 Optuna [97]	41
Rumus 2.10 Optuna pada CatBoost [101].....	43
Rumus 2.11 SHAP [110].....	45
Rumus 2.12 Akurasi [115]	48
Rumus 2.13 Presisi [115]	48
Rumus 2.14 Recall [115].....	49
Rumus 2.15 F1 Score [115].....	49



DAFTAR LAMPIRAN

Lampiran A Turnitin Similarity Report	198
Lampiran B Form Konsultasi Bimbingan.....	202

