



SKRIPSI

**KLASIFIKASI TINGKAT KEPARAHAN
KECELAKAAN LALU LINTAS BERBASIS
CATBOOST PADA DATA YANG TIDAK
SEIMBANG MENGGUNAKAN SMOTENC DAN
OPTUNA**

MASLAHATUL KAUNAINI AYATILAH
NPM 21081010261

DOSEN PEMBIMBING

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
Achmad Junaidi, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABA**

LEMBAR PENGESAHAN

KLASIFIKASI TINGKAT KEPARAHAN KECELAKAAN LALU LINTAS BERBASIS CATBOOST PADA DATA YANG TIDAK SEIMBANG MENGUNAKAN SMOTENC DAN OPTUNA

Oleh :
MASLAHATUL KAUNAINI AYATILAH
NPM. 21081010261

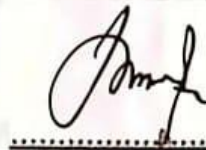
Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 11 Desember 2025

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
NPT. 222198 60 816400



(Pembimbing I)

Achmad Junaldi, S.Kom., M.Kom.
NIP. 19781110 202521 1 048



(Pembimbing II)

Fetty Tri Anggraeny, S.Kom., M.Kom.
NIP. 19820211 2021212 005



(Penguji I)

Firza Prima Aditiawan, S.Kom., MTI.
NIP. 19860523 202121 1 003



(Penguji II)

Mengetahui,
Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

**KLASIFIKASI TINGKAT KEPARAHAN KECELAKAAN LALU LINTAS
BERBASIS CATBOOST PADA DATA YANG TIDAK SEIMBANG
MENGUNAKAN SMOTENC DAN OPTUNA**

Oleh :
MASLAHATUL KAUNAINI AYATILAH
NPM. 21081010261



Menyetujui,
Koordinator Program Studi Informatika
Fakultas Ilmu Komputer

Fetty Tri Anggraeny, S.Kom., M.Kom.
NIP. 19820211 2021212 005

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertandatangan di bawah ini:

Nama : Maslahatul Kaunaini Ayatilah
NPM : 21081010261
Program : Sarjana (S1)
Program Studi : Informatika
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 15 Januari 2026
Yang Membuat Pernyataan,




Maslahatul Kaunaini Ayatilah
NPM. 21081010261

ABSTRAK

Nama Mahasiswa / NPM : Maslahatul Kaunaini Ayatilah / 21081010261
Judul Skripsi : Klasifikasi Tingkat Keparahan Kecelakaan Lalu Lintas Berbasis Catboost pada Data yang Tidak Seimbang Menggunakan Smotenc dan Optuna
Dosen Pembimbing : 1. Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
2. Achmad Junaidi, S.Kom., M.Kom.

Kecelakaan lalu lintas merupakan permasalahan kompleks di Kota Surabaya yang berdampak signifikan terhadap keselamatan publik dan kerugian sosio-ekonomi. Salah satu tantangan utama dalam pemodelan klasifikasi tingkat keparahan kecelakaan adalah adanya fenomena ketidakseimbangan data (*imbalanced data*), di mana jumlah kejadian luka ringan jauh lebih dominan dibandingkan kategori luka berat atau meninggal dunia. Penelitian ini bertujuan untuk mengklasifikasikan tingkat keparahan kecelakaan menjadi tiga kelas, yaitu Kelas 0 (Ringan), Kelas 1 (Sedang), dan Kelas 2 (Berat) dengan mengimplementasikan algoritma CatBoost. Untuk mengatasi kendala distribusi kelas yang tidak merata, diterapkan metode *Synthetic Minority Over-sampling Technique for Nominal and Continuous* (SMOTENC), sementara optimasi *hyperparameter* dilakukan menggunakan kerangka kerja Optuna. Evaluasi model dilakukan menggunakan metode *hold-out* dengan proporsi 60% data latih dan 40% data uji. Hasil eksperimen menunjukkan bahwa integrasi SMOTENC dan Optuna secara signifikan meningkatkan performa prediktif model. Model terbaik, yakni CatBoost-SMOTENC-Optuna, menghasilkan akurasi sebesar 97,20%, presisi 95%, *recall* 90%, dan *F1-Score* sebesar 92%. Performa ini menunjukkan peningkatan stabilitas yang substansial dibandingkan model CatBoost standar yang hanya menghasilkan *F1-Score* sebesar 32%. Temuan penelitian ini mengonfirmasi bahwa kombinasi SMOTENC dan Optuna efektif dalam meningkatkan sensitivitas model terhadap kelas minoritas (Sedang dan Berat), sehingga dapat memberikan kontribusi bagi pihak berwenang dalam merumuskan strategi mitigasi risiko kecelakaan yang lebih akurat dan terukur.

Kata kunci : Kecelakaan Lalu Lintas, CatBoost, SMOTENC, Optuna, Ketidakseimbangan Data.

ABSTRACT

Nama Mahasiswa / NPM : Maslahatul Kaunaini Ayatilah / 21081010261
Judul Skripsi : Classification of Traffic Accident Severity Levels
Based on CatBoost on Imbalanced Data Using
SMOTENC and Optuna
Dosen Pembimbing : 1. Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
2. Achmad Junaidi, S.Kom., M.Kom.

Traffic accidents represent a complex issue in Surabaya, having a significant impact on public safety and socio-economic loss. A primary challenge in accident severity classification modeling is the phenomenon of data imbalance, where slight injury cases are considerably more dominant than severe or fatal categories. This study aims to classify accident severity into three classes: Class 0 (Slight), Class 1 (Moderate), and Class 2 (Severe) by implementing the CatBoost algorithm. To address the uneven class distribution, the Synthetic Minority Over-sampling Technique for Nominal and Continuous (SMOTENC) was applied, while hyperparameter optimization was conducted using the Optuna framework. Model evaluation utilized the hold-out method with a 60% training and 40% testing data split. Experimental results demonstrate that the integration of SMOTENC and Optuna significantly enhances the model's predictive performance. The CatBoost-SMOTENC-Optuna model achieved superior performance with an accuracy of 97.20%, precision of 95%, recall of 90%, and an F1-Score of 92%. Overall, this research proves that the combination of SMOTENC and Optuna optimization is effective in handling imbalanced data and increasing model sensitivity across all severity classes. These findings are expected to serve as a scientific reference for authorities in formulating more accurate risk mitigation strategies and traffic safety policies.

Keywords: Road Accidents, CatBoost, SMOTENC, Optuna, Data Imbalance.

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat, hidayah, dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Klasifikasi Tingkat Keparahannya Kecelakaan Lalu Lintas Berbasis Catboost pada Data yang Tidak Seimbang Menggunakan Smotenc dan Optuna”** dapat terselesaikan dengan baik. Penulis menyampaikan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu baik secara langsung maupun tidak langsung selama proses penyusunan skripsi ini. Secara khusus, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, MMT., IPU. selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom., M.Kom. selaku Koordinator Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Bapak Hendra Maulana, S.Kom, M.Kom., selaku Dosen Wali, yang senantiasa mendampingi dan memberikan arahan bijak selama penulis menempuh pendidikan di perguruan tinggi ini.
5. Ibu Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT., selaku Dosen Pembimbing I, atas segala dedikasi, kesabaran, serta transfer ilmu yang luar biasa dalam mengarahkan penulis hingga penelitian ini mencapai standar yang diharapkan.
6. Bapak Achmad Junaidi, S.Kom., M.Kom., selaku Dosen Pembimbing II, atas setiap masukan kritis, bimbingan teknis, serta motivasi yang diberikan selama proses penyusunan karya ini.
7. Ibu Fetty Tri Anggraeny, S.Kom., M.Kom., selaku Penguji Pertama, atas arahan serta saran-saran membangun yang diberikan guna penyempurnaan skripsi ini.
8. Bapak Firza Prima Aditiawan, S.Kom., MTI., selaku Dosen Penguji II, yang telah memberikan masukan, saran, serta memberikan perspektif berharga sehingga penulis dapat memperbaiki penelitian ini.
9. Seluruh Dosen Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur atas segala ilmu pengetahuan yang diberikan kepada penulis selama masa perkuliahan. Semoga

penulis dapat mengamalkan ilmu pengetahuan yang telah diberikan dan akan menjadi bekal yang berguna bagi penulis di masa yang akan datang.

10. Ayah, Ibu, dan Adik, sebagai pilar kekuatan utama. Terima kasih atas doa yang tidak pernah putus, pengorbanan materi, serta kasih sayang tak terhingga yang menjadi motivasi terbesar bagi penulis untuk menyelesaikan studi ini dengan baik.
11. Sahabat dan Teman-Teman Dekat, yang telah hadir sebagai sistem pendukung terbaik, selalu membantu penulis dan terlibat dalam setiap proses pengerjaan skripsi ini. Terima kasih atas bantuan, tawa, dan semangat yang kalian berikan di saat-saat tersulit selama proses penulisan ini.
12. Rekan-Rekan Perjuangan Informatika Angkatan 2021, atas kebersamaan, rasa solidaritas, dan memori tak terlupakan selama menjalani dinamika perkuliahan bersama.
13. Apresiasi untuk diri sendiri, atas kegigihan, ketangguhan, dan kerelaan untuk tidak menyerah dalam menghadapi setiap tantangan akademik hingga titik terakhir.

Semoga Tuhan Yang Maha Esa memberikan balasan yang setimpal kepada semua pihak yang telah memberikan bantuan, bimbingan maupun motivasi yang bermanfaat bagi penulis. Penulis menyadari bahwa karya ini masih memiliki ruang untuk perbaikan. Oleh karena itu, setiap masukan dan kritik yang bersifat konstruktif sangat penulis harapkan. Semoga skripsi ini mampu memberikan kontribusi positif bagi pengembangan ilmu pengetahuan dan bermanfaat bagi banyak pihak.

Surabaya, 15 Januari 2026

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	i
LEMBAR PERSETUJUAN	ii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xiii
DAFTAR KODE PROGRAM	xiv
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terdahulu	6
2.2 Kecelakaan Lalu Lintas.....	10
2.3 <i>Machine Learning</i>	12
2.4 Klasifikasi	13
2.5 Catboost (<i>Categorical Boosting</i>)	14
2.6 Split Dataset	16
2.7 SMOTENC.....	18
2.8 OPTUNA	20
2.9 Evaluasi Model.....	22
2.9.1 <i>Confusion Matrix</i>	23
BAB III METODOLOGI PENELITIAN	26
3.1 Tahapan Penelitian	26
3.2 Studi Literatur	27

3.3 Pengumpulan Data	27
3.3.1 Sumber Data.....	27
3.3.2 Atribut Data.....	27
3.4 <i>Preprocessing Data</i>	29
3.4.1 Pembersihan Data (<i>Data Cleaning</i>)	29
3.4.2 Transformasi Fitur (<i>Feature Transformation</i>)	29
3.4.3 <i>Feature Engineering</i>	31
3.5 Modeling	33
3.5.1 Split Data.....	33
3.5.2 Penyeimbang Data (SMOTENC).....	34
3.5.3 Catboost	36
3.5.4 <i>Hyperparameter</i> OPTUNA.....	39
3.6 Evaluasi	41
3.6.1 Metrik Evaluasi Klasifikasi.....	41
3.6.2 Skenario Pengujian.....	44
3.6.3 Implementasi Dashboard.....	45
BAB IV HASIL DAN PEMBAHASAN	46
4.1 Pengumpulan Data	46
4.2 <i>Preprocessing Data</i>	47
4.2.1 Pembersihan Data.....	47
4.2.2 Ekstraksi dan Transformasi Fitur	48
4.2.3 <i>Feature Engineering</i>	50
4.3 Perancangan Model.....	52
4.3.1 Split Data.....	52
4.3.2 SMOTENC.....	53
4.3.3 Modelling Catboost.....	54
4.4 Skenario Pengujian.....	55
4.4.1 Proporsi Data.....	55
4.4.2 <i>Learning rate</i>	61
4.4.3 Iteration	68
4.4.4 <i>Depth</i>	75
4.4.5 Hasil Keseluruhan Pengujian	81
4.5 Evaluasi Model.....	84

4.5.1 Model Catboost	85
4.5.2 Model Catboost – SMOTENC	87
4.5.3 Model Catboost – SMOTENC – Optuna	89
4.5.4 Analisis Performa Model	90
4.6 Implementasi Dashboard.....	93
BAB V KESIMPULAN DAN SARAN	107
5.1 Kesimpulan	107
5.2 Saran.....	108
DAFTAR PUSTAKA.....	109
LAMPIRAN.....	113

DAFTAR GAMBAR

Gambar 2. 1 Kecelakaan Lalu Lintas di Jalan Ahmad Yani, Surabaya	10
Gambar 2. 2 Ilustrasi <i>Supervised Learning</i>	12
Gambar 2. 3 Kelas pada Klasifikasi.....	13
Gambar 2. 4 Mekanisme Catboost dalam Pembentukan Model	14
Gambar 2. 5 Ilustrasi <i>Train-Test Split</i>	16
Gambar 2. 6 Ilustrasi SMOTENC	18
Gambar 2. 7 Alur Optimasi <i>Hyperparameter</i> dengan Optuna.....	20
Gambar 3. 1 <i>Flowchart</i> Alur Penelitian.....	26
Gambar 3. 2 <i>Preprocessing Data</i>	29
Gambar 3. 3 Tahap Modeling	33
Gambar 3. 4 Distribusi data sebelum SMOTENC	35
Gambar 3. 5 Distribusi Data setelah SMOTENC	35
Gambar 4. 1 Tampilan 5 baris awal dataset	47
Gambar 4. 2 Penanganan <i>Missing Values</i>	48
Gambar 4. 3 Data After <i>Preprocessing</i>	48
Gambar 4. 4 Transformasi Fitur Waktu	49
Gambar 4. 5 Hasil Fitur Kendaraan	50
Gambar 4. 6 Label Target	52
Gambar 4. 7 Hasil Pembagian Data Latih dan Data Uji	53
Gambar 4. 8 <i>Confusion Matrix</i> 80:20	56
Gambar 4. 9 <i>Classification Report</i>	57
Gambar 4. 10 <i>Confusion Matrix</i> 70:30	58
Gambar 4. 11 <i>Classification Report</i> 70:30.....	58
Gambar 4. 12 <i>Confusion Matrix</i> 60:40	59
Gambar 4. 13 <i>Classification Report</i> 60:40.....	60
Gambar 4. 14 <i>Confusion Matrix</i> LR 0.01-0.1	62
Gambar 4. 15 <i>Classification Report</i> LR 0.01-0.1	63
Gambar 4. 16 <i>Confusion Matrix</i> LR 0.1-0.2	64
Gambar 4. 17 <i>Classification Report</i> LR 0.1-0.2	64
Gambar 4. 18 <i>Confusion Matrix</i> LR 0.2-0.3	65
Gambar 4. 19 <i>Classification Report</i> LR 0.2-0.3	66

Gambar 4. 20 <i>Confusion Matrix</i> Iterasi 100-400	69
Gambar 4. 21 <i>Classification Report</i> Iterasi 100-400	70
Gambar 4. 22 <i>Confusion Matrix</i> Iterasi 400-700	71
Gambar 4. 23 <i>Classification Report</i> Iterasi 400-700	71
Gambar 4. 24 <i>Confusion Marix</i> Iterasi 700-1000	73
Gambar 4. 25 <i>Classification Report</i> Iterasi 400-1000	73
Gambar 4. 26 <i>Confusion Matrix Depth</i> 4-6	76
Gambar 4. 27 <i>Classification Report Depth</i> 4-6.....	77
Gambar 4. 28 <i>Confusion Matrix Depth</i> 7-8	78
Gambar 4. 29 <i>Classification Report Depth</i> 7-8.....	78
Gambar 4. 30 <i>Confusion Matrix Depth</i> 9-10	79
Gambar 4. 31 <i>Classification Repot Depth</i> 9-10	80
Gambar 4. 32 <i>Confusion Matrix</i> Model Catboost.....	85
Gambar 4. 33 <i>Classification Report</i> Model Catboost	86
Gambar 4. 34 <i>Confusion Matrix</i> Model Catboost–SMOTENC.....	87
Gambar 4. 35 <i>Classification Report</i> Model Catboost–SMOTENC	88
Gambar 4. 36 <i>Confusion Matrix</i> Model Catboost–SMOTENC–Optuna.....	89
Gambar 4. 37 <i>Classification Report</i> Model Catboost–SMOTENC–Optuna	90
Gambar 4. 38 Dashboard.....	93
Gambar 4. 39. Proses Awal.....	94
Gambar 4. 40. Jumlah Kecelakaan per Tahun (2020-2024)	95
Gambar 4. 41. Tren Kecelakaan per Bulan - Tahun 2020	96
Gambar 4. 42. Tren Kecelakaan per Bulan - Tahun 2021	97
Gambar 4. 43. Tren Kecelakaan per Bulan - Tahun 2022	98
Gambar 4. 44. Tren Kecelakaan per Bulan - Tahun 2023	99
Gambar 4. 45. Tren Kecelakaan per Bulan - Tahun 2024	100
Gambar 4. 46. Distribusi Hari Kejadian	100
Gambar 4. 47. Distribusi Jam Kejadian	101
Gambar 4. 48. Proporsi Kecelakaan Berdasarkan Hari.....	102
Gambar 4. 49. Tingkat Keparahan Kecelakaan per Tahun	103
Gambar 4. 50. Frekuensi Jenis Kendaraan Terlibat	104
Gambar 4. 51. Lokasi Jumlah Kecelakaan Terbanyak.....	105

DAFTAR TABEL

Tabel 3. 1 Atribut Data.....	27
Tabel 3. 2 Atribut Waktu Kejadian pada Dataset	30
Tabel 3. 3 Atribut Waktu Kejadian setelah Ekstraksi Fitur	30
Tabel 3. 4 Kategori Hari.....	30
Tabel 3. 5 Kategori Jam	30
Tabel 3. 6 Menambahkan Fitur total_korban	31
Tabel 3. 7 Menambahkan Fitur total_kendaraan.....	31
Tabel 3. 8 Kategori Jenis Kendaraan	31
Tabel 3. 9 Nilai AEK	32
Tabel 3. 10 Ruang Pencarian <i>Hyperparameter</i>	40
Tabel 3. 11 Skenario Pengujian	44
Tabel 4 1. Hasil Pengujian Split Data	61
Tabel 4 2. Hasil Pengujian Skenario <i>Learning rate</i>	67
Tabel 4 3. Hasil Pengujian Skenario Iteration	74
Tabel 4 4. Hasil Pengujian Skenario <i>Depth</i>	81
Tabel 4 5. Hasil Uji Skenario	82
Tabel 4 6. Parameter Terbaik	84
Tabel 4 7. Analisis Performa Model	91

DAFTAR KODE PROGRAM

Kode Program 4. 1 Read Dataset	46
Kode Program 4. 2 Data Cleaning	47
Kode Program 4. 3 Ekstraksi Waktu Kejadian	48
Kode Program 4. 4 Klasifikasi Hari	49
Kode Program 4. 5 Ekstraksi Jenis Kendaraan Terlibat	49
Kode Program 4. 6 Fitur Total Kendaraan.....	50
Kode Program 4. 7 Label Keparahan Tingkat Kecelakaan.....	51
Kode Program 4. 8 Split data	52
Kode Program 4. 9 SMOTENC	53
Kode Program 4. 10 Model Catboost.....	54
Kode Program 4. 11 Implementasi <i>hyperparameter</i>	54