

**KLASIFIKASI TINGKAT KECELAKAAN LANTAS POLDA
JATIM MENGGUNAKAN KNN, SVM, DECISION TREE, DAN
NAÏVE BAYES DENGAN SMOTE**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Skripsi

Sultan Adyatma Rangga Setiawan

0000061918

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

2025

**KLASIFIKASI TINGKAT KECELAKAAN LANTAS POLDA
JATIM MENGGUNAKAN KNN, SVM, DECISION TREE, DAN
NAÏVE BAYES DENGAN SMOTE**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Skripsi

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Sistem Komputer

Sultan Adyatma Rangga Setiawan

0000061918

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Sultan Adyatma Rangga Setiawan

Nomor Induk Mahasiswa : 00000061918

Program studi : Sistem Informasi

Skripsi dengan judul:

KLASIFIKASI TINGKAT KECELAKAAN LANTAS POLDA JATIM MENGGUNAKAN KNN, SVM, DECISION TREE, DAN NAÏVE BAYES DENGAN SMOTE merupakan hasil karya saya sendiri bukan plagiat dari karya 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 skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 29 November 2023

A handwritten signature in black ink is written over a yellow rectangular stamp. The stamp contains the text 'SULTAN ADYATMA RANGGA SETIAWAN' and a unique identification number '54545AJX017204510'.

(Sultan Adyatma Rangga Setiawan)

U M I N
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

HALAMAN PENGESAHAN

Sign.com Document ID: 107b5e9df6 - Page 1/1

HALAMAN PENGESAHAN

Skripsi dengan judul

"Klasifikasi Tingkat Kecelakaan Lantas Polda Jatim Menggunakan SVM, KNN, Decision Tree, dan Naive Bayes dengan SMOTE"

Oleh

Nama : Sultan Adyatma Rangga Setiawan
NIM : 00000061918
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Kamis, 05 Juni 2025

Pukul 13.30 s.d 15.30 dan dinyatakan

LULUS

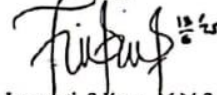
Dengan susunan penguji sebagai berikut.

Ketua Sidang



Ririn Ikana Desanti, S.Kom., M.Kom.
031305880001

Penguji



Dr. Irmawati, S.Kom., M.M.S.I.
0805097703

Pembimbing



Prof. Dr. Friska Natalia, S.Kom., M.T.
0306128307

Ketua Program Studi Sistem Informasi



Ririn Ikana Desanti, S.Kom., M.Kom.
031305880001

iii

Klasifikasi Data Kecelakaan ..., Sultan Adyatma Rangga Setiawan, Universitas Multimedia Nusantara

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Sultan Adyatma Rangga Setiawan

NIM : 00000061918

Program Studi : Sistem Informasi

Fakultas : Teknik & Informatika

Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*) atas skripsi saya yang berjudul.

KLASIFIKASI TINGKAT KECELAKAAN LANTAS POLDA JATIM MENGGUNAKAN KNN, SVM, DECISION TREE, DAN NAÏVE BAYES DENGAN SMOTE.

Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Multimedia Nusantara berhak menyimpan, mengalihmediakan/mengalihformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 31 Januari 2025

Yang menyatakan,



(Sultan Adyatma Rangga Setiawan)

KATA PENGANTAR

Segala puji syukur kepada Tuhan yang telah memberikan rahmat dan kemudahan kepada penulis untuk dapat menyelesaikan laporan skripsi dengan judul “KLASIFIKASI TINGKAT KECELAKAAN LANTAS POLDA JATIM MENGGUNAKAN KNN, SVM, DECISION TREE, DAN NAÏVE BAYES DENGAN SMOTE” secara tepat waktu. Penulis menyadari betapa berdampak bimbingan dan dukungan dalam menuntaskan skripsi ini. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Dr. Andrey Andoko, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dewan 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. Prof. Friska Natalia, S.Kom., M.T., sebagai Pembimbing 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.

Harapan sebesar besarnya agar karya ilmiah ini dapat bermanfaat dan membantu pembaca sebagai sumber informasi atau wawasan untuk penelitian dan karya ilmiah selanjutnya.

Tangerang, 31 Januari 2025


(Sultan Adyatma Rangga Setiawan)

KLASIFIKASI TINGKAT KECELAKAAN LANTAS POLDA JATIM MENGGUNAKAN KNN, SVM, DECISION TREE, DAN NAÏVE BAYES DENGAN SMOTE

Sultan Adyatma Rangga Setiawan

ABSTRAK

Kecelakaan lalu lintas merupakan salah satu permasalahan utama dalam sistem transportasi, yang berdampak signifikan terhadap keselamatan pengguna jalan, baik pengendara maupun pejalan kaki. Di Provinsi Jawa Timur, jumlah pengguna kendaraan bermotor terus mengalami peningkatan setiap tahunnya. Berdasarkan data gabungan dari BPSJ Jatim dan GoodStats, tercatat sebanyak 23.591.000 pengguna pada tahun 2022, meningkat menjadi 25.201.000 pada 2023, dan mencapai 26.527.000 pada tahun 2024. Seiring dengan pertumbuhan ini, angka kecelakaan lalu lintas juga meningkat, dengan 32.000 kasus tercatat oleh Polda Jawa Timur pada tahun 2023, menjadikan provinsi ini sebagai wilayah dengan jumlah kecelakaan tertinggi di Indonesia pada tahun tersebut.

Tingginya angka kecelakaan dipengaruhi oleh berbagai faktor, baik dari sisi perilaku pengendara seperti ketidakpatuhan terhadap aturan lalu lintas, maupun dari faktor eksternal seperti waktu padat kendaraan, kondisi cuaca, dan infrastruktur jalan yang tidak memadai. Sebagai instansi yang menjaga keamanan masyarakat, Polda perlu melakukan analisis untuk mengurangi peningkatan angka dan tingkat kecelakaan di sebuah wilayah.

Penelitian ini bertujuan untuk mengklasifikasikan tingkat kecelakaan lalu lintas di Jawa Timur menggunakan algoritma SVM, KNN, Decision Tree, dan Naïve Bayes. Untuk mengatasi ketidakseimbangan data, diterapkan metode Synthetic Minority Over-sampling Technique (SMOTE). Model terbaik akan dikembangkan menjadi aplikasi sederhana agar dapat mengklasifikasi tingkat kecelakaan sebagai pendukung pemilihan keputusan oleh pihak Polda Jatim dan membantu keefisienan kinerja pendataan instansi. Pendekatan ini diharapkan dapat memberikan kontribusi terhadap upaya pencegahan dan pengendalian kecelakaan lalu lintas secara lebih efektif di wilayah Jawa Timur.

Kata kunci: Decision Tree, Jawa Timur, Klasifikasi Tingkat Kecelakaan, KNN, Lalu lintas, SVM, SMOTE

CLASSIFICATION OF TRAFFIC ACCIDENT LEVEL IN EAST JAVA POLDA USING KNN, SVM, DECISION TREE, AND NAÏVE BAYES WITH SMOTE

Sultan Adyatma Rangga Setiawan

ABSTRACT (English)

Traffic accidents are one of the main issues in the transportation system, significantly impacting the safety of road users, both drivers and pedestrians. In East Java Province, the number of motor vehicle users continues to increase every year. Based on combined data from BPSP Jatim and GoodStats, there were 23,591,000 users in 2022, increasing to 25,201,000 in 2023, and reaching 26,527,000 in 2024. Along with this growth, the number of traffic accidents also increased, with 32,000 cases recorded by the East Java Police in 2023, making this province the area with the highest number of accidents in Indonesia that year.

The high number of accidents is influenced by various factors, both from the side of driver behavior such as non-compliance with traffic rules, and from external factors such as peak vehicle hours, weather conditions, and inadequate road infrastructure. As an institution that maintains public safety, the police need to conduct an analysis to reduce the increase in the number and rate of accidents in an area.

This research aims to classify the level of traffic accidents in East Java using SVM, KNN, Decision Tree, and Naïve Bayes algorithms. To address data imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) method is applied. The best model will be developed into a simple application to classify accident levels as a decision-making support tool for the East Java Regional Police and to enhance the efficiency of the agency's data collection performance. This approach is expected to contribute to more effective traffic accident prevention and control efforts in the East Java region.

Keywords: *Accident Rate Classification East Java, Decision Tree, KNN, SMOTE, SVM, Traffic*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iii
KATA PENGANTAR	v
ABSTRAK.....	vi
<i>ABSTRACT (English)</i>	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xiii
DAFTAR RUMUS	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan dan Manfaat Penelitian	5
1.4.1 Tujuan Penelitian	5
1.4.2 Manfaat Penelitian	5
1.5 Sistematika Penulisan	6
BAB II LANDASAN TEORI.....	7
2.1 Penelitian Terdahulu	7
2.2 Teori Penelitian	12
2.2.1 Laka Lantas	12
2.2.2 Metode <i>oversampling</i> SMOTE	13
2.3 Framework dan Algoritma Penelitian	13
2.3.1 SVM.....	14
2.3.2 K-Nearest Neighbors (KNN)	15
2.3.3 Decision Tree	16
2.3.4 Naïve Bayes	17
2.3.5 CRISP-DM.....	18

2.3.6	Confusion Matrix	18
2.4	Tools Penelitian.....	19
2.4.1	Python	19
2.4.2	Google Colab	20
2.4.3	Visual Studio Code	20
BAB III METODOLOGI PENELITIAN		21
3.1	Objek Penelitian	21
3.2	Metode Penelitian.....	21
3.2.1	Alur Penelitian	21
3.2.2	Metode Data Mining	23
3.2.3	Metode Pengolahan Data	26
3.3	Teknik Pengumpulan Data	27
3.4	Variabel Penelitian	27
3.4.1	Variabel Tergantung (Dependent Variable).....	27
3.4.2	Variabel Bebas (Independent Variable).....	27
3.5	Teknik Analisis Data.....	27
BAB IV ANALISIS DAN HASIL PENELITIAN		30
4.1	Implementasi.....	30
4.1.1	Business Understanding	30
4.1.2	Data Understanding.....	31
4.1.3	Data Preparation.....	36
4.1.4	Data Modeling.....	45
4.1.5	Evaluation	52
4.1.6	Deployment.....	59
4.2	Perbandingan.....	69
4.3	Diskusi	71
BAB V SIMPULAN DAN SARAN.....		74
5.1	Simpulan	74
5.2	Saran.....	75
DAFTAR PUSTAKA		76
LAMPIRAN.....		80

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	7
Tabel 3.1 Perbandingan <i>framework data mining</i>	23
Tabel 3.2 Perbandingan Algoritma Pengolahan Data	26
Tabel 3.3 Perbandingan <i>Software Analisis Data</i>	28
Tabel 4.1 Rincian Distribusi Tingkat Kecelakaan	35
Tabel 4.2 Hasil Grid Search Model SVM	52
Tabel 4.3 Hasil Grid Search Model KNN	52
Tabel 4.4 Hasil Grid Search Model Decision Tree	53
Tabel 4.5 Hasil Grid Search Model Naïve Bayes	54
Tabel 4.6 Rincian Hasil <i>Confusion Matrix</i> Seluruh Model	58
Tabel 4.7 Perbandingan Model Terhadap Penelitian Terdahulu	69



DAFTAR GAMBAR

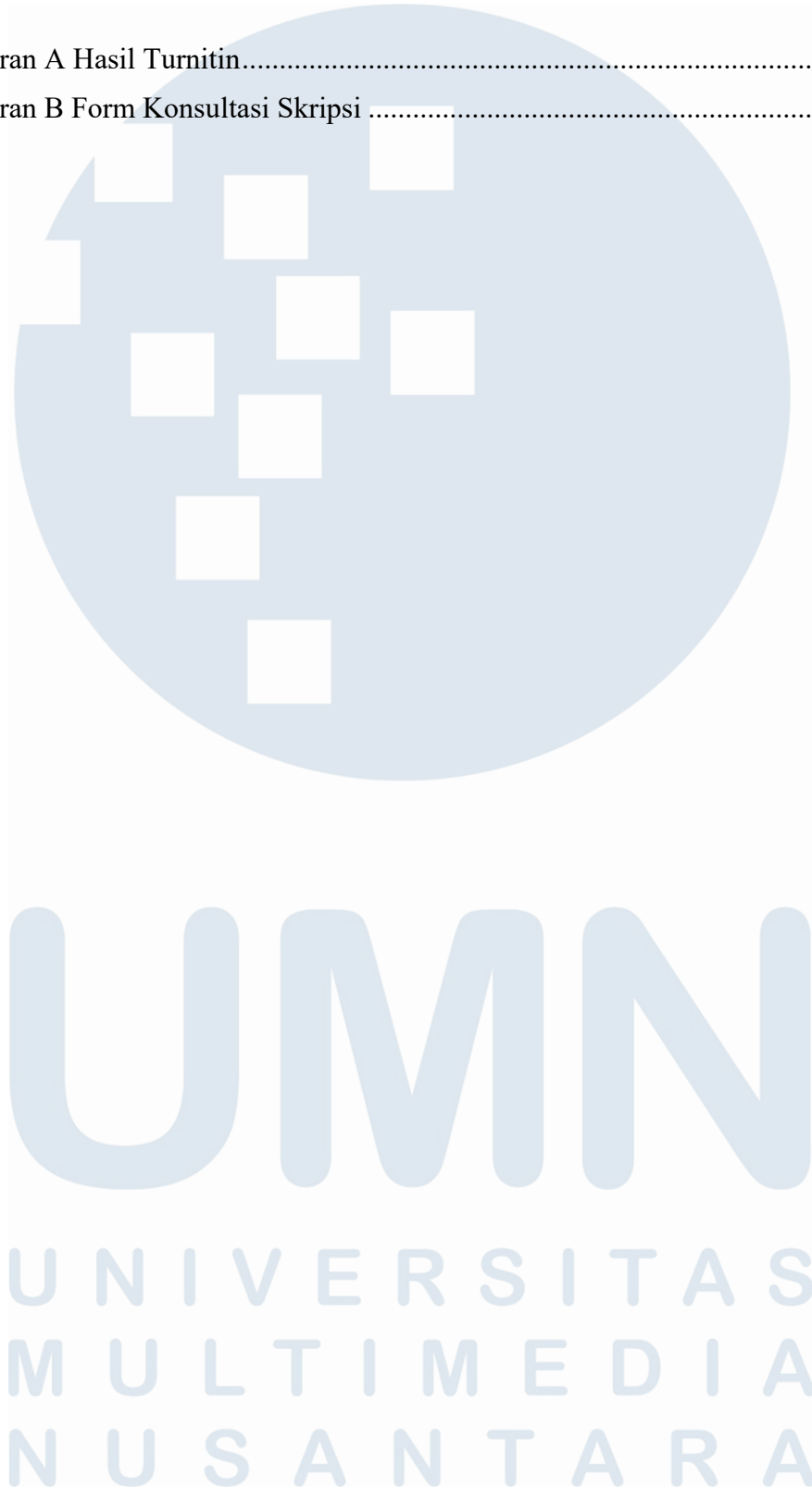
Gambar 1.1 Peningkatan Pengguna Kendaraan Bermotor 2022-2024 [1], [2].....	1
Gambar 1.2 Data Laporan Kecelakaan Polda 2023[3].....	2
Gambar 2.1 Proses <i>Sampling</i> SMOTE[17].....	13
Gambar 2.2 Proses SVM Linear[18].....	14
Gambar 2.4 Proses SVM Non-Linear[19].....	15
Gambar 2.3 Proses KNN[20].....	16
Gambar 2.5 Proses Decision Tree[21].....	17
Gambar 3.1 Diagram Alur Penelitian.....	22
Gambar 4.1 Fitur pada Dataset.....	32
Gambar 4.2 Frekuensi Faktor Utama.....	34
Gambar 4.3 Distribusi Tingkat Kecelakaan.....	35
Gambar 4.4 Import Data dari Google Drive.....	36
Gambar 4.5 Import Package.....	37
Gambar 4.6 Import Dataset.....	37
Gambar 4.7 Tampilan Dataset.....	38
Gambar 4.8 Menampilkan Heatmap Korelasi Fitur.....	38
Gambar 4.9 Heatmap Korelasi Fitur.....	39
Gambar 4.10 Proses Labelling.....	40
Gambar 4.11 Variabel Dependen dan Independen.....	41
Gambar 4.12 Data Latih dan Uji.....	41
Gambar 4.13 Menampilkan Distribusi Tanpa SMOTE.....	42
Gambar 4.14 Data Latih Tanpa SMOTE.....	42
Gambar 4.15 Menampilkan Distribusi dengan SMOTE.....	43
Gambar 4.16 Data Latih dengan SMOTE.....	43
Gambar 4.17 Menampilkan Distribusi dengan <i>Random Under Sampling</i>	44
Gambar 4.18 Data Latih dengan <i>Random Under Sampling</i>	44
Gambar 4.19 Menampilkan Perbandingan <i>Hyperparameter</i> Model.....	45
Gambar 4.20 Hasil Ringkasan Model Berdasarkan Akurasi Data Uji.....	46
Gambar 4.21 Percobaan Menggunakan <i>Random Under Sampling</i> SVM.....	47
Gambar 4.22 Hasil Model <i>Random Under Sampling</i> SVM.....	48

Gambar 4.23 Membentuk Rasio Pembagian Data 90:10	49
Gambar 4.24 Hasil Model Tanpa <i>Sampling</i> dengan Rasio 90:10	50
Gambar 4.25 <i>Feature Importance</i> untuk Decision Tree	51
Gambar 4.26 Menampilkan <i>Confusion Matrix</i> Model Tanpa <i>Sampling</i>	55
Gambar 4.27 Perbandingan <i>Confusion Matrix</i> Model Tanpa <i>Sampling</i>	56
Gambar 4.28 Menampilkan <i>Confusion Matrix</i> Seluruh Model	57
Gambar 4.29 Instal <i>library</i> untuk Visual Studio Code	59
Gambar 4.30 Membentuk File <i>Pickle</i> Setiap Model.....	60
Gambar 4.31 Membentuk <i>app.py</i>	61
Gambar 4.32 <i>Run</i> Streamlit <i>app.py</i>	62
Gambar 4.33 Prediksi Tingkat Kecelakaan Ringan	63
Gambar 4.34 Prediksi Tingkat Kecelakaan Waspada	64
Gambar 4.35 Prediksi Tingkat Kecelakaan Darurat	65
Gambar 4.36 Membentuk Otomasi <i>Real-Time</i> untuk Data Berkelanjutan	66
Gambar 4.37 Aplikasi Web Pelatihan Model Terhadap Data Berkelanjutan	67
Gambar 4.38 Hasil Pengembangan Model Menggunakan Data Kecelakaan 2024	68
Gambar 4.39 Hasil Pengembangan Model dan Data	69



DAFTAR LAMPIRAN

Lampiran A Hasil Turnitin.....	80
Lampiran B Form Konsultasi Skripsi	81



DAFTAR RUMUS

Rumus 2. 1 Rumus Accuracy[24].	18
Rumus 2. 2 Rumus Precision[24].....	19
Rumus 2. 3 Rumus Recall[24]	19
Rumus 2. 4 Rumus F1-Score[24].....	19

