



**PENERAPAN KLASIFIKASI MENGGUNAKAN ALGORITMA
ARTIFICIAL NEURAL NETWORK (ANN) DENGAN *FEATURE
SELECTION RANDOM FOREST* UNTUK MENGETAHUI
FAKTOR-FAKTOR PENYEBAB KECELAKAAN LALU
LINTAS**

SKRIPSI

DARNISA AZZAHRA NASUTION

1610511077

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAKARTA**

**FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA**

2020



**PENERAPAN KLASIFIKASI MENGGUNAKAN ALGORITMA
ARTIFICIAL NEURAL NETWORK (ANN) DENGAN *FEATURE
SELECTION* RANDOM FOREST UNTUK MENGETAHUI FAKTOR-
FAKTOR PENYEBAB KECELAKAAN LALU LINTAS**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Komputer**

DARNISA AZZAHRA NASUTION

1610511077

UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2020

PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Darnisa Azzahra Nasution

NIM : 1610511077

Tanggal : 17 Mei 2020

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 17 Mei 2020

Yang Menyatakan,



(Darnisa Azzahra Nasution)

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional “Veteran” Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Darnisa Azzahra Nasution

NIM :1610511077

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional “Veteran” Jakarta Hak Bebas Royalti Non eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

Penerapan Klasifikasi Menggunakan Algoritma Artificial Neural Network (Ann) Dengan Feature Selection Random Forest Untuk Mengetahui Faktor-Faktor Penyebab Kecelakaan Lalu Lintas

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada Tanggal : 17 Mei 2020
Yang Menyatakan,



(Darnisa Azzahra Nasution)

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Skripsi berikut:

Nama : Darnisa Azzahra Nasution
NIM : 1610511077
Program Studi : S1 Informatika
Judul Skripsi : Penerapan Klasifikasi Menggunakan Algoritma Artificial Neural Network (Ann) Dengan Feature Selection Random Forest Untuk Mengetahui Faktor-Faktor Penyebab Kecelakaan Lalu Lintas

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Yuni Widiastiwi, S.Kom., M.Si

Penguji 1



Bambang Tri Wahyono, S.Kom., M.Si

Penguji 2



Iin Ernawati, S.Kom., M.Si.

Pembimbing I



Nurul Chamidah, S.Kom., M.Kom.

Pembimbing II



Dr. Ermatita, M.Kom.
Dekan



Anita Muliawati, S.Kom., MTI.

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 18 Juni 2020



ABSTRAK

Kecelakaan Lalu Lintas adalah sebuah kejadian yang tidak terduga, kemudian melibatkan pengendara, kendaraan dengan pengguna jalan lain maupun tidak, sehingga mengakibatkan kerugian harta benda hingga korban manusia. Saat ini kecelakaan lalu lintas menjadi penyulut kematian yang cukup besar di Indonesia, sehingga perlu adanya tindakan-tindakan yang mampu mengurangi bahkan mencegah terjadinya kecelakaan. Penelitian ini dilakukan untuk menerapkan klasifikasi algoritma *Artificial Neural Network* (ANN) dengan *Random Forest* sebagai teknik untuk *feature selection* dalam mengetahui faktor-faktor penyebab dari kecelakaan lalu lintas berdasarkan tingkat kecelakaan, yang nantinya dapat membantu memberikan informasi kepada masyarakat. Hasil yang diperoleh adalah klasifikasi dengan *feature selection* menghasilkan akurasi 75.17%, yang sedikit meningkat dibandingkan tanpa adanya *features selection*, yaitu 75.14%. Faktor-faktor penyebab kecelakaan lalu lintas teratas berdasarkan pengujian yang telah dilakukan adalah koordinat GPS-lintang dan koordinat GPS-bujur.

Kata Kunci : Kecelakaan Lalu Lintas, *Random Forest*, ANN

ABSTRACT

Traffic accident is an unexpected event, then involving motorists, vehicles with other road users or not, resulting in loss of property to human victims. Currently, traffic accidents are a significant source of death in Indonesia, so actions are needed to reduce or even prevent accidents. This research was conducted to apply the classification of Artificial Neural Network (ANN) algorithm with Random Forest as a technique for feature selection in knowing the causes of traffic accidents based on accident rates, which can later help provide information to the public. The results obtained are the classification with feature selection produces an accuracy of 75.17%, which slightly increases compared to the absence of feature selection, which is 75.14%. The factors that cause top traffic accidents based on tests that have been done are GPS-latitude coordinates and GPS-longitude coordinates.

Keywords : Traffic Accident, Random Forest, ANN

KATA PENGANTAR

Puji dan syukur atas kehadiran Allah SWT atas segala rahmat dan karunia-Nya, shalawat serta salam tak lupa kita haturkan kepada Nabi Muhammad SAW, keluarga serta sahabatnya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul **“Penerapan Klasifikasi Menggunakan Algoritma *Artificial Neural Network (Ann)* Dengan *Feature Selection Random Forest* untuk Mengetahui Faktor-Faktor Penyebab Kecelakaan Lalu Lintas”**.

Penulisan Tugas Akhir ini merupakan salah satu syarat untuk memperoleh gelar sarjana Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jakarta. Rasa terimakasih penulis ucapkan kepada:

1. Kepada orang tua dan keluarga yang saya cintai, yang telah memberikan dukungan dan doa kepada penulis selama masa pengerjaan skripsi ini.
2. Ibu Dr. Ermatita, M.Kom., selaku Dekan Fakultas Ilmu Komputer
3. Ibu Anita Muliawati, S.Kom., MTI., selaku Kepala Program Studi Informatika.
4. Ibu Iin Ernawati, S.Kom., M.Si., dan Ibu Nurul Chamidah, S.Kom., M.Kom., selaku dosen pembimbing I & II skripsi yang membantu penulis dalam penyusunan skripsi sehingga dapat menyelesaikan skripsi dengan baik.
5. Ibu Mayanda Mega Santoni, S.Kom., M.Kom., selaku dosen pembimbing akademik.
6. Ibu, Bapak Dosen Program Studi Informatika UPN “Veteran” Jakarta atas ilmu-ilmu yang bermanfaat.
7. Kepada Pihak Korlantas Polri, yang telah memberikan izin untuk mengambil data Kecelakaan Lalu Lintas.
8. Kepada teman-teman penulis mahasiswa Program Studi Informatika Angkatan 2016 yang tidak dapat penulis sebutkan namanya satu persatu.

Akhir kata, semoga skripsi ini dapat bermanfaat bagi para pembacanya.

Jakarta, 18 Mei 2020

Penulis

Darnisa Azzahra Nasution

DAFTAR ISI

PERNYATAAN ORISINALITAS	iii
LEMBAR PENGESAHAN	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR SIMBOL	xiv
BAB 1	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Ruang Lingkup	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.6 Luaran yang Diharapkan	3
1.7 Sistematika Penulisan	3
BAB 2	5
TINJAUAN PUSTAKA	5
2.1 Kecelakaan Lalu Lintas	5
2.2 <i>Data Mining</i>	5
2.2.1 Definisi <i>Data Mining</i>	7
2.2.2 Teknik <i>Data Mining</i>	7
2.2.3 Tahapan <i>Data Mining</i>	8
2.3 <i>Imbalanced Data</i>	10
2.4 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	11
2.5 Klasifikasi	11
2.6 <i>Feature Selection</i>	12
2.7 <i>Random Forest</i>	13
2.7.1 Karakteristik <i>Random Forest</i> adalah sebagai berikut.	14
2.7.2 Penjelasan algoritma <i>Random Forests</i>	14

2.8 Artificial Neural Network (ANN).....	17
2.8.1 Jenis Jenis ANN.....	18
2.8.2 Fungsi Aktivasi.....	19
2.8.3 Metode <i>Feed-forward Backpropagation</i>	21
2.9 Evaluasi Kinerja Pengklasifikasian	25
2.9 Studi Literatur.....	26
BAB 3	29
METODOLOGI PENELITIAN.....	29
3.1 Kerangka Pikir.....	29
3.1.1 Identifikasi Masalah.....	30
3.1.2 Studi Literatur	30
3.1.3 Pengumpulan Data	30
3.1.4 <i>Preprocessing Data</i>	30
3.1.5 Pembagian Data	31
3.1.6 Pemodelan Data	31
3.1.7 Evaluasi.....	31
3.2 Perangkat Penelitian	31
3.3 Tempat Penelitian.....	32
3.4 Jadwal Penelitian	32
BAB 4	33
HASIL DAN PEMBAHASAN.....	33
4.1 Data	33
4.2 Praproses Data	38
4.2.1 <i>Data Cleaning</i>	38
4.2.2 <i>Feature Selection</i>	58
4.3 Pembagian Data.....	63
4.3.1 Data Latih (<i>Training</i>).....	63
4.3.2 Data Uji (<i>Testing</i>)	64
4.4 Artificial Neural Network (ANN).....	65
4.4.1 Bobot dan Bias.....	65
4.4.2 Fungsi Aktifasi.....	65
4.4.3 Maksimum <i>Epoch</i>	66
4.4.4 Jumlah Neuron pada Lapisan Tersembunyi (<i>Hidden layer</i>).....	67
4.4.5 <i>Learning rate</i> (Laju Pembelajaran).....	67

4.4.6 Hasil Pelatihan ANN	70
4.4.7 Hasil Pengujian ANN	72
4.4.8 Hasil Klasifikasi dengan <i>Features Selection</i>	74
4.4.9 Faktor-Faktor Penyebab Kecelakaan Lalu Lintas.....	85
BAB 5	87
PENUTUP	87
5.1 Kesimpulan.....	87
5.2 Saran.....	87
DAFTAR PUSTAKA	89
RIWAYAT HIDUP.....	92
LAMPIRAN.....	93

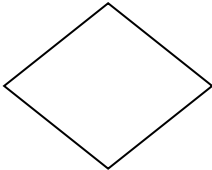
DAFTAR GAMBAR

Gambar 1. Proses Pencarian Pengetahuan	6
Gambar 2. <i>Flowchart Random Forest</i>	15
Gambar 3. Model Arsitektur <i>Neural Network</i>	17
Gambar 4. <i>Single-Layer Neural Network</i>	19
Gambar 5. <i>Multilayer Neural Network</i>	19
Gambar 6. Grafik Fungsi Sigmoid	20
Gambar 7. Arsitektur Jaringan <i>Feed Forward Backpropagation</i>	21
Gambar 8. Kerangka Pikir.....	29
Gambar 9. Data Kecelakaan Lalu Lintas (2016-2019)	33
Gambar 10. Data Setelah Penghapusan Row	39
Gambar 11. Jumlah <i>Missing Value</i> Tiap <i>Features</i>	40
Gambar 12. Data Setelah Penghapusan Kolom	41
Gambar 13. Jumlah <i>Missing Value</i> Tiap <i>Features</i>	42
Gambar 14. Jumlah <i>Missing Value</i> Setelah Imputasi	43
Gambar 15. Data dengan <i>Missing Value</i>	43
Gambar 16. Data Setelah Pengisian <i>Missing Value</i>	44
Gambar 17. Data <i>Input</i> Setelah Penghapusan <i>Features</i> Tidak Relevan	45
Gambar 18. Transformasi <i>Feature</i> Tanggal Kejadian	46
Gambar 19. Data Setelah Label <i>Encoding</i>	49
Gambar 20. Data untuk Kelas Rugi Materi Saja Sebelum <i>Oversampling</i>	52
Gambar 21. Data Kelas Rugi Materi Saja Setelah <i>Oversampling</i>	53
Gambar 22. Data Setelah <i>Oversampling</i>	54
Gambar 23. Data Setelah Normalisasi	58
Gambar 24. Salah Satu <i>Tree</i> dari <i>Random Forest</i>	61
Gambar 25. Grafik <i>Features Importance</i>	63
Gambar 26. Data Latih (<i>Training</i>)	64
Gambar 27. Data Uji (<i>Testing</i>).....	65
Gambar 28. Arsitektur Pemodelan ANN	70

DAFTAR TABEL

Table 1. Jadwal Penelitian.....	32
Tabel 2. Pembagian Kategori Waktu	47
Tabel 3. <i>Label Encoding</i> Waktu Kejadian	47
Tabel 4. <i>Dummy Encoding</i> Waktu Kejadian.....	48
Tabel 5. <i>Label Encoding</i> Tingkat Kecelakaan	50
Tabel 6. <i>Label To Categorical</i> Tingkat Kecelakaan	50
Table 7. Jumlah Data Tiap Kelas Sebelum <i>Oversampling</i>	51
Table 8. Jumlah Data Tiap Kelas Setelah <i>Oversampling</i>	52
Tabel 9. Error OOB dari Pengujian Mtry.....	59
Tabel 10. Error OOB dari Pengujian Jumlah Pohon.....	59
Tabel 11. <i>Feature Importance</i>	62
Tabel 12. Percobaan Variasi Jumlah <i>Epoch</i>	66
Tabel 13. Variasi Parameter Pelatihan	67
Tabel 14. Hasil Pelatihan ANN.....	71
Tabel 15. Hasil Pengujian ANN	72
Tabel 16. Hasil Percobaan 17 <i>Features</i>	75
Tabel 17. Hasil Percobaan 16 <i>Features</i>	76
Tabel 18. Hasil Percobaan 15 <i>Features</i>	76
Tabel 19. Hasil Percobaan 14 <i>Features</i>	77
Tabel 20. Hasil Percobaan 13 <i>Features</i>	78
Tabel 21. Hasil Percobaan 12 <i>Features</i>	79
Tabel 22. Hasil Percobaan 11 <i>Features</i>	80
Tabel 23. Hasil Percobaan 10 <i>Features</i>	80
Tabel 24. Hasil Percobaan 9 <i>Features</i>	81
Tabel 25. Hasil Percobaan 8 <i>Features</i>	82
Tabel 26. Hasil Percobaan 7 <i>Features</i>	83
Tabel 27. Hasil Percobaan 6 <i>Features</i>	84
Table 28 Perbandingan Akurasi Klasifikasi.....	85

DAFTAR SIMBOL

Simbol	Nama Simbol	Keterangan
	Simbol Proses	Menggambarkan Proses
	Simbol Decision	Simbol pemilihan proses berdasarkan kondisi yang ada
	Simbol arah data atau arus data	Sebagai petunjuk arah data dan arus data pada proses
	Simbol Terminator	Simbol untuk permulaan atau akhir dari suatu kegiatan
	Simbol Input-Output	Simbol yang menyatakan proses input dan output tanpa tergantung jenis peralatannya