

Analisis Kecelakaan Lalu Lintas Berbasis Decision Tree Terintegrasi Forward Selection dan Metode ADASYN

Muhammad Wildan Hadinata¹⁾, Taghfirul Azhima Yoga Siswa^{2*)}, Wawan Joko Pranoto³⁾

^{1,2,3)} Sains Dan Teknologi, Universitas Muhammadiyah Kalimantan Timur
Jl. Ir. H. Juanda No. 15, Sidodadi, Kec. Samarinda Ulu, Kota Samarinda, Kalimantan Timur 75124
Email : 2111102441108@umkt.ac.id¹⁾, tay758@umkt.ac.id²⁾, wjp337@umkt.ac.id³⁾

Abstrak

Kecelakaan lalu lintas di Indonesia, khususnya di Provinsi Kalimantan Timur, menunjukkan tren yang mengkhawatirkan, dengan tercatat 665 kasus pada tahun 2023. Penelitian ini bertujuan untuk mengembangkan model klasifikasi tingkat keparahan kecelakaan lalu lintas dengan mengintegrasikan metode *Adaptive Synthetic Sampling* (ADASYN) dan *Forward Selection* ke dalam algoritma *Decision Tree*. Data yang digunakan berasal dari Polresta Kota Samarinda periode 2020-2024, mencakup 35 fitur. Proses penelitian melibatkan tahapan data *preprocessing*, seleksi fitur menggunakan *Forward Selection*, penyeimbangan data menggunakan ADASYN, dan pembagian data dengan teknik *10-Fold Cross-Validation*. Evaluasi model dilakukan menggunakan metrik *confusion matrix*, yang mencakup *akurasi*, *presisi*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa fitur yang paling berpengaruh terhadap klasifikasi adalah cuaca, jumlah korban meninggal dunia, luka berat, dan luka ringan. Penggunaan ADASYN meningkatkan akurasi dari 69,44% menjadi 76,83%, dan penerapan *Forward Selection* lebih lanjut menaikkan akurasi hingga 99,06%. Namun, nilai akurasi yang sangat tinggi ini masih perlu diuji lebih lanjut untuk memastikan generalisasi model terhadap data di luar sampel. Penelitian ini dibatasi pada penggunaan satu algoritma klasifikasi tanpa perbandingan dengan algoritma lain. Temuan ini menunjukkan bahwa kombinasi ADASYN dan *Forward Selection* dapat meningkatkan performa model klasifikasi pada data kecelakaan lalu lintas yang tidak seimbang.

Kata Kunci : *Kecelakaan Lalu Lintas; Decision Tree; Forward Selection; ADASYN; Klasifikasi*

Abstract

Traffic accidents in Indonesia, particularly in East Kalimantan Province, have shown an alarming trend, with 665 cases recorded in 2023. This study aims to develop a classification model for traffic accident severity by integrating the *Adaptive Synthetic Sampling* (ADASYN) and *Forward Selection* methods into the *Decision Tree* algorithm. The dataset was obtained from the Samarinda City Police Department for the period 2020–2024, comprising 35 features. The research process involved several stages, including data preprocessing, feature selection using *Forward Selection*, data balancing using ADASYN, and data partitioning with the *10-Fold Cross-Validation* technique. The model was evaluated using confusion matrix metrics, including accuracy, precision, recall, and F1-score. The results showed that the most influential features for classification were weather conditions, number of fatalities, serious injuries, and minor injuries. The application of ADASYN improved the model's accuracy from 69.44% to 76.83%, while the implementation of *Forward Selection* further increased the accuracy to 99.06%. However, this exceptionally high accuracy still requires further validation to ensure the model's generalizability to unseen data. This study is limited to the use of a single classification algorithm without comparison to other models. These findings indicate that the combination of ADASYN and *Forward Selection* can significantly enhance classification performance on imbalanced traffic accident data.

Keywords: *Traffic Accidents; Decision Tree; Forward Selection; ADASYN; Classification*

1. PENDAHULUAN

Organisasi Kesehatan Dunia (WHO) melaporkan bahwa setiap tahun, sekitar 1,3 juta orang meninggal akibat kecelakaan lalu lintas, menjadikannya salah satu penyebab utama tingginya kematian di dunia. Di Indonesia, angka kecelakaan lalu lintas juga menunjukkan tren yang mengkhawatirkan. Pada tahun 2023, terjadi 665 kasus kecelakaan lalu lintas di provinsi Kalimantan Timur [1]. Secara lebih spesifik, Kota Samarinda mengalami peningkatan jumlah kecelakaan lalu lintas. Pada tahun 2022 ke 2023, terjadi peningkatan kecelakaan lalu lintas dengan tercatat 150 kasus kecelakaan dengan 75 korban meninggal [2]. Angka ini meningkat pada tahun 2024 menjadi 233 kasus, dengan 67 korban meninggal dunia [3]. Peningkatan ini menunjukkan urgensi untuk memahami faktor-faktor yang memengaruhi kecelakaan lalu lintas di Samarinda. Kondisi tersebut dapat merugikan banyak pihak, jika tidak diidentifikasi dengan tepat akan mengakibatkan dampak yang fatal.

Analisis klasifikasi kecelakaan lalu lintas memiliki peran penting dalam menurunkan jumlah kecelakaan dan meningkatkan keselamatan jalan [4]. Salah satu pendekatan yang dapat digunakan adalah *machine learning*, karena menyediakan berbagai teknik dan algoritma untuk menganalisis data secara efisien. Penelitian [5], menunjukkan bahwa metode *transfer learning* dari model *machine learning* dapat digunakan untuk mengidentifikasi dan mengklasifikasikan cedera dan kondisi jalan dengan akurasi 89% pada dataset *traffignetV1* dan 79% pada dataset *wound type*. Sementara itu, [6] menggunakan *regresi logistik multinomial* untuk mengukur akurasi model prediksi serta mengidentifikasi variabel yang mempengaruhi kecelakaan lalu lintas, dengan akurasi 72%. Sedangkan, penelitian [7] menggunakan model *machine learning* untuk membuat sistem deteksi kecelakaan dengan metode *deep learning* mencapai akurasi tinggi sebesar 93,93%, dan akurasi validasi sebesar 90,80%, yang menunjukkan model tersebut layak diterapkan.

Penerapan metode *machine learning*, khususnya melalui algoritma *Decision tree*, memiliki peran krusial dalam menganalisis dan mengklasifikasikan data kecelakaan lalu lintas. *Decision Tree* merupakan algoritma *machine learning* yang sangat populer, karena kemampuannya dalam memodelkan serta memahami hubungan kompleks antara variabel-variabel yang terlibat dalam suatu proses pengambilan keputusan [8]. Berbagai penelitian terdahulu seperti [9] menunjukkan bahwa *Decision*

Tree mampu mengklasifikasi tingkat kecelakaan lalu lintas dengan akurasi mencapai 93%. Sementara itu, [10] menggunakan algoritma C4.5, yang merupakan salah satu varian dari *Decision Tree*, untuk klasifikasi status gizi balita dan berhasil mencapai akurasi sempurna sebesar 100% menggunakan bantuan *software RapidMiner*.

Seringkali, proses klasifikasi memiliki banyak fitur yang dapat memengaruhi hasil prediksi. Oleh karena itu, pemilihan fitur yang tepat memiliki peran penting, salah satu metodenya adalah *Forward Selection*. Metode tersebut yang akan membantu menghasilkan model yang lebih efisien dan akurat. *Forward Selection* adalah metode seleksi fitur dalam *machine learning* yang digunakan untuk menentukan subset fitur yang paling relevan dari suatu dataset [11]. Penelitian yang dilakukan oleh [12] membuktikan efektivitas metode *Forward Selection* dalam meningkatkan akurasi model klasifikasi. Dalam studi tersebut, metode ini digunakan untuk memilih atribut yang lebih sedikit namun tetap relevan yang mampu meningkatkan akurasi dari 69,90% menjadi 79,11%. Hasil tersebut menunjukkan bahwa *Forward Selection* dapat membantu dalam mengidentifikasi dan memilih fitur yang paling signifikan, sehingga dapat meningkatkan kinerja model klasifikasi secara signifikan. Dalam penelitian yang dilakukan [13], menunjukkan metode *Forward Selection* memiliki peran dalam memilih fitur-fitur yang paling relevan dan signifikan yang berkontribusi terhadap hasil klasifikasi yang membantu mengurangi dimensi data dan meningkatkan efisiensi algoritma. Hasil utama dari penelitian tersebut, penerapan metode *Forward Selection* terbukti dapat meningkatkan akurasi sebesar 94,32%. Sedangkan penelitian yang dilakukan [14], menjelaskan proses seleksi *Forward Selection* menggunakan model tanpa fitur dan secara bertahap menambahkan fitur satu per satu berdasarkan peningkatan performa model, seperti akurasi dan nilai R^2 . Hasil evaluasi menunjukkan bahwa kombinasi *K-Nearest Neighbor* (KNN) dengan *Forward Selection* mencapai akurasi, presisi, *recall*, dan skor F1 sebesar 100%. Berdasarkan penelitian tersebut, proses seleksi dimulai dengan satu fitur, kemudian secara bertahap menambahkan fitur-fitur lain yang dapat meningkatkan kinerja model pada setiap tahap. Proses ini berlanjut sampai tidak ada lagi peningkatan kinerja yang berarti atau hingga jumlah fitur yang diinginkan tercapai.

Dalam konteks klasifikasi data kecelakaan lalu lintas, jumlah kelas non-kecelakaan sering lebih banyak dibandingkan kelas kecelakaan, yang

menciptakan Ketidakseimbangan dalam klasifikasi. Ketidakseimbangan data ini merupakan masalah yang sering dihadapi dalam klasifikasi, di mana jumlah sampel dalam satu kelas jauh sedikit dibandingkan dengan jumlah sampel di kelas lainnya [15]. Untuk mengatasi masalah ini, penggunaan metode *ADASYN* (*Adaptive Synthetic Sampling*), diharapkan menghasilkan data sintetis dari kelas minoritas, meningkatkan representasi data kecelakaan dan kinerja model klasifikasi [16]. Penelitian [17] menunjukkan bahwa *ADASYN* berhasil mengatasi ketidakseimbangan dataset, meskipun metode SMOTE lebih unggul dengan akurasi 81% tanpa *oversampling*. Metode *ADASYN* diharapkan meningkatkan kinerja model dengan memberikan data yang lebih seimbang dan mencakup lebih banyak contoh dari kelas minoritas. Penelitian lain oleh [18] juga menunjukkan penerapan *ADASYN* dalam klasifikasi biner untuk mendeteksi serangan *malware* dengan akurasi 99.86%. Dengan demikian, *ADASYN* tidak hanya mengatasi ketidakseimbangan data, tetapi juga meningkatkan akurasi dan efektivitas model dalam memprediksi kecelakaan lalu lintas.

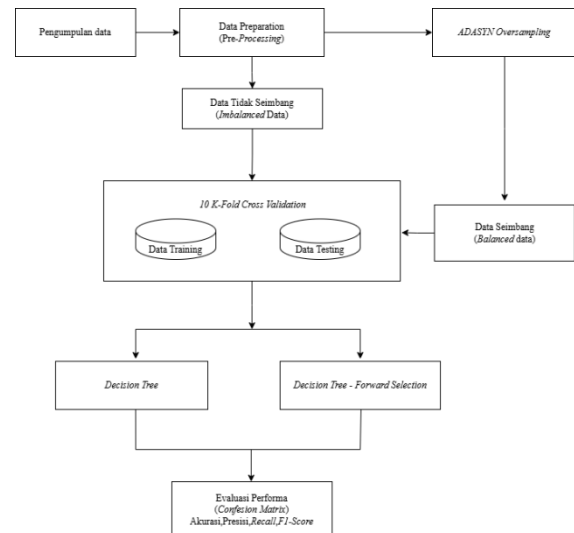
Berdasarkan uraian latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah bagaimana pengaruh metode *ADASYN* dalam mengatasi ketidakseimbangan data kecelakaan lalu lintas dan apakah metode *Forward Selection* dapat meningkatkan akurasi model *Decision Tree* dalam klasifikasi kecelakaan.

Penelitian ini bertujuan untuk menganalisis pengaruh metode *ADASYN* dalam mengatasi ketidakseimbangan data pada klasifikasi kecelakaan lalu lintas dan mengevaluasi efektivitas metode *Forward Selection* dalam meningkatkan akurasi algoritma *Decision Tree* untuk klasifikasi tingkat keparahan kecelakaan lalu lintas.

2. METODE PENELITIAN

2.1 Prosedur Penelitian

Dengan alur penelitian ini menggambarkan klasifikasi data kecelakaan lalu lintas menggunakan *Forward Selection*, *ADASYN*, dan *Decision Tree*.



Gambar 1. Alur Penelitian Klasifikasi Kecelakaan Lalu Lintas

Pada Gambar 1. diperlihatkan alur penelitian klasifikasi kecelakaan lalu lintas yang dimulai dari pengumpulan data, dilanjutkan dengan tahap *pre-processing*, identifikasi ketidakseimbangan kelas, dan penanganannya menggunakan metode *ADASYN oversampling* untuk menghasilkan data yang seimbang. Data kemudian dibagi menggunakan teknik *10 K-Fold Cross Validation* menjadi data latih dan data uji, yang selanjutnya digunakan untuk membangun model klasifikasi menggunakan algoritma *Decision Tree* dan *Decision Tree* dengan pendekatan *Forward Selection*. Terakhir, performa model dievaluasi menggunakan *Confusion Matrix* dan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-Score*.

2.3 Pengumpulan Data

Pada penelitian ini menggunakan data Laka Lantas yang bersifat privasi yang didapatkan secara langsung di kantor Polresta Kota Samarinda. Kantor Polresta Kota Samarinda dengan izin akses yang sudah disetujui. Pada periode tahun 2020-2024. Dari data instansi tersebut mendapatkan fitur-fitur yang digunakan meliputi nomor laka, polres, tanggal kejadian, waktu kejadian, tanggal dilaporkan, waktu dilaporkan, tingkat kecelakaan, jumlah meninggal, jumlah luka berat, jumlah luka ringan, koordinat gps – lintang, koordinat gps – bujur, titik acuan atau referensi, jarak ke lokasi kecelakaan, arah dari titik acuan ke lokasi kecelakaan, informasi khusus, tipe kecelakaan, kondisi cahaya, cuaca, kecelakaan menonjol, nomor jalan, nama jalan, titik jalan, fungsi jalan, kelas jalan, tipe jalan, bentuk geometri, kondisi permukaan jalan, batas kecepatan dilokasi, kemiringan jalan, status jalan, penyebab perkiraan nilai rugi material kendaraan, total nilai rugi material

non kendaraan dan keterangan kerugian. Lokasi penelitian di Polresta Kota Samarinda Jl. Slamet Riyadi No.1, Karang Asam Ulu, Kec. Sungai Kunjang, Kota Samarinda, Kalimantan Timur 75126.

2.4 Data Pre-Processing

a. Data Selection

Data Selection yaitu memilih data yang relevan untuk analisis.

b. Data Cleaning

Pada tahap data cleaning, membersihkan data dari kesalahan, duplikasi atau nilai yang kosong.

c. Data Transformation

Proses transformasi meliputi konversi variabel kategorikal (misalnya cuaca, jenis jalan) ke dalam bentuk numerik menggunakan label encoding.

d. Data Balancing

Data Balancing merupakan teknik untuk menyeimbangkan data agar jumlah sampel setiap kelas lebih merata.

2.5 Pembagian Data

Dataset dibagi menjadi data latih dan data uji agar performa model bisa diukur secara objektif. Penelitian ini menggunakan metode 10-Fold Cross-Validation.

iterasi 1	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
iterasi 2	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
iterasi 3	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
iterasi 4	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
iterasi 5	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
iterasi 6	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
iterasi 7	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
iterasi 8	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
iterasi 9	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
iterasi 10	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10

Gambar 2. Proses 10-Fold Cross-Validation

Pada gambar 2. ditampilkan skema 10-fold cross validation, di mana setiap baris mewakili satu iterasi. Blok berwarna merah menunjukkan bagian data yang digunakan sebagai data uji pada iterasi tersebut, sedangkan blok berwarna putih menunjukkan data latih. Proses ini dilakukan sebanyak 10 kali (10 fold), sehingga setiap bagian data memiliki kesempatan menjadi data uji sekali dan data latih sembilan kali. Hasil evaluasi dari masing-masing fold kemudian dihitung rata-ratanya untuk memberikan estimasi performa model yang lebih akurat.

2.6 Permodelan Decision Tree

Algoritma Decision Tree digunakan dalam penelitian ini karena mampu menangani data numerik maupun kategorikal, serta memberikan interpretasi yang jelas dalam bentuk pohon keputusan. Pada tahap ini, model dibentuk dengan cara memisahkan data berdasarkan fitur-fitur yang memberikan information gain tertinggi.

```
# Model Decision Tree yang ditingkatkan
clf = DecisionTreeClassifier(
    max_depth=3,
    min_samples_leaf=5,
    class_weight='balanced',
    criterion='entropy',
    random_state=42
)

# K-Fold
cv = StratifiedKFold(n_splits=10, shuffle=True, random_state=42)

# Simpan metrik
conf_matrices = []
accuracies = []
```

Gambar 3. Kode Program Decision Tree

Model Decision TreeClassifier dikonfigurasi dengan parameter max_depth=3, min_samples_leaf=5, class_weight='balanced', criterion='entropy', dan random_state=42 guna mengurangi risiko overfitting serta memastikan hasil yang konsisten. Evaluasi dilakukan menggunakan StratifiedKFold dengan 10 lipatan, shuffle=True, dan random_state=42 agar distribusi kelas pada setiap lipatan tetap seimbang. Pada tiap iterasi, model dilatih dan diuji, kemudian hasil evaluasi berupa confusion matrix dan akurasi dicatat untuk dianalisis secara keseluruhan.

2.7 Permodelan Decision Tree Dengan Forward Selection

Untuk meningkatkan efisiensi dan akurasi model, penelitian ini mengintegrasikan teknik seleksi fitur Forward Selection ke dalam proses pelatihan Decision Tree. Metode ini dimulai dengan model kosong, kemudian fitur ditambahkan satu per satu berdasarkan kontribusinya terhadap

```
# Gunakan Forward Selection
sfs = SequentialFeatureSelector(
    dt,
    n_features_to_select="auto", # bisa diganti angka jika ingin fix jumlah fitur
    direction="forward",
    scoring="accuracy",
    cv=StratifiedKFold(n_splits=5),
    n_jobs=-1
)

# Fit selector ke data
sfs.fit(X, y)

# Ambil fitur terpilih
selected_features = x.columns[sfs.get_support()]

# Buat DataFrame hasil
df_selected = pd.DataFrame({'Selected Features': selected_features})
```

peningkatan akurasi model.

Gambar 4. Tampilan Kode Forward Selection

SequentialFeatureSelector dikonfigurasi menggunakan *DecisionTreeClassifier* sebagai estimator, dengan metode seleksi fitur maju (*forward selection*) secara otomatis berdasarkan skor akurasi, menggunakan validasi silang stratified 5-fold, serta pemrosesan paralel (*n_jobs=-1*). Proses seleksi dilakukan melalui *fit(x, y)*, lalu hasil dari *get_support()* digunakan untuk mengekstrak nama-nama kolom pada *x* yang terpilih, dan disajikan dalam bentuk *DataFrame* yang menampilkan fitur-fitur terbaik.

2.8 Evaluasi Model

Kinerja model dievaluasi menggunakan *confusion matrix*, yang memvisualisasikan hasil klasifikasi berdasarkan jumlah data yang benar atau salah diklasifikasi.

$$ACCURACY = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

Keterangan :

- TP = *True Positives* (Prediksi benar bahwa kecelakaan terjadi)
- TN = *True Negatives* (Prediksi benar bahwa tidak terjadi kecelakaan)
- FP = *False Positives* (Prediksi salah bahwa kecelakaan terjadi padahal tidak)
- FN = *False Negatives* (Prediksi salah bahwa tidak terjadi kecelakaan padahal terjadi)

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pra-pemrosesan data

Penelitian ini menggunakan data Laka Lantas yang bersifat privasi yang didapatkan secara langsung di kantor Polresta Kota Samarinda (2020-2024). Terdapat 35 fitur. Melalui pemilihan fitur yang terpilih yaitu 13 fitur relevan yang melambangkan waktu, lokasi, kondisi lingkungan dan dampak kecelakaan.

tingkat kecelakaan	jumlah meninggal dunia	jumlah luka berat	jumlah luka ringan	cuaca	rumah	jenis jalan	nomor polisi kendaraan	jenis kendaraan	jenis kecelakaan	jenis korban
132	Ringan	0	0	1	Cerah	Adiri	Bah	40.0	Mancala	
13	Sedang	0	1	0	Cerah	Adiri	Bah	50.0	Mancala	
14	Ringan	0	0	2	Cerah	Adiri	Bah	50.0	Mancala	
24	Ringan	0	0	2	Cerah	Adiri	Bah	50.0	Mancala	
31	Ringan	0	0	2	Cerah	Adiri	Bah	40.0	Mancala	
...	...	...	...	...	...	...	...	...	...	...
1649	Ringan	0	0	1	Cerah	Kabutar	Bah	50.0	Mancala	
1327	Ringan	0	0	1	Cerah	Kabutar	Bah	40.0	Mancala	
1684	Ringan	1	0	0	Cerah	Kabutar	Bah	50.0	Mancala	
957	Sedang	0	1	0	Cerah	Kabutar	Bah	50.0	Mancala	
1691	Ringan	1	0	0	Cerah	Kabutar	Bah	50.0	Mancala	

Gambar 1. Data Selection

Setelah data dibersihkan dari nilai yang hilang dan duplikasi, tersisa data sebanyak 1.167 record. Data kategorikal seperti cuaca, fungsi jalan dan tingkat keparahan kemudian ditransformasikan menjadi numerik menggunakan teknik *label encoding*. Untuk mengatasi ketidakseimbangan pada kelas (data ringan/sedang 911 dan berat 148),

digunakan metode *ADASYN* sehingga penyebarannya menjadi seimbang (911 pada data ringan/sedang dan 912 pada berat).

3.2 Hasil Permodelan Decision Tree

Model Decision Tree diuji menggunakan *10K-Fold Cross Validation* untuk menilai kinerja awal. Hasil evaluasi menunjukkan bahwa model mampu memprediksi tingkat keparahan kecelakaan secara akurat, dengan rata – rata akurasi sebesar 66,44%.Tingkat akurasi ini belum tergolong tinggi, pada tingkat akurasi yang sedang pada model ini memiliki kemampuan dasar yang cukup baik dalam mengklasifikasikan data.

Tabel 2. Hasil pengujian model Decision Tree dan Rata – rata nya

Fold	Akurasi	Precision	Recall	F1 - Score
1	67,21%	80,49%	67,21%	71,80%
2	50,82%	78,30%	50,82%	57,87%
3	73,63%	75,28%	73,63%	74,43%
4	80,77%	80,44%	80,77%	80,60%
5	80,22%	79,54%	80,22%	79,87%
6	76,37%	77,51%	76,37%	76,93%
7	86,81%	84,30%	86,81%	83,68%
8	81,32%	80,71%	81,32%	81,00%
9	48,90%	77,70%	48,90%	56,02%
10	48,35%	75,84%	48,35%	55,70%
Rata - Rata	69,44%	79,01%	69,44%	71,79%

Hasil evaluasi tabel 2. pada model *Decision Tree* tanpa menerapkan *ADASYN* dan *Forward Selection* menunjukkan hasil kinerja rata-rata model *Decision Tree* berdasarkan pengujian menggunakan 10 fold validasi. Akurasi rata -rata yang dicapai oleh model adalah 69,44%, menunjukkan bahwa hampir 70% dari data diklasifikasikan dengan benar. Pada akurasi rata – rata *precision* yaitu 79,01% menunjukkan bahwa model secara akurat mengidentifikasi data positif. Pada *recall* hasil rata – rata nya adalah 69,44%. Ini menunjukkan bahwa persentase data positif yang berhasil diakui oleh model masih sesuai dengan tingkat akurasi keseluruhan. Nilai skor F1 rata - rata 71,79% mencerminkan keseimbangan antara akurasi dan penarikan, memberikan citra umum yang dilakukan oleh model pohon keputusan dengan sangat baik, tetapi telah membaik, terutama ketika berhadapan dengan data yang mungkin tidak seimbang.

3.3 Hasil Permodelan Decision Tree setelah ADA SYN

Pengujian awal dilakukan menggunakan model *Decision Tree* tanpa *Forward Selection*. Evaluasi model dengan *10K-Fold Cross Validation* menunjukkan akurasi rata-rata sebesar 76.83%.

Tingkat akurasi yang tinggi ini menunjukkan kemampuan model untuk mengklasifikasi data secara akurat, dengan tingkat kesalahan yang rendah.

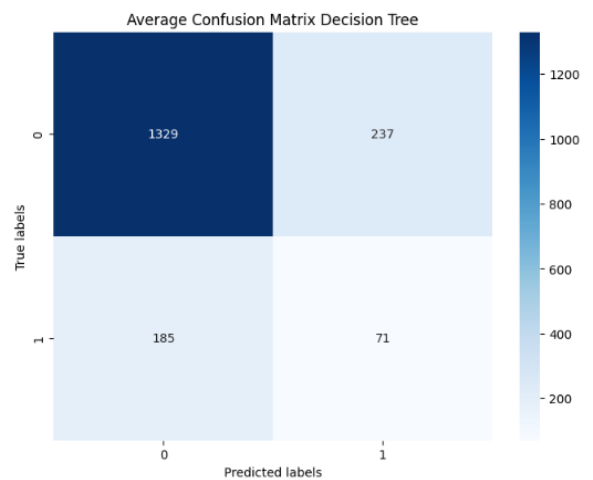
Tabel 3. Hasil Pengujian Model *Decision Tree* setelah *ADASYN*

<i>Fold</i>	Akurasi	<i>Precision</i>	<i>Recall</i>	<i>F1 - Score</i>
1	79.23%	81.14%	79.23%	80.11%
2	81.42%	79.53%	81.42%	80.38%
3	73.63%	75.28%	73.63%	74.43%
4	80.77%	80.44%	80.77%	80.60%
5	80.22%	79.54%	80.22%	79.87%
6	76.37%	77.51%	76.37%	76.93%
7	82.42%	79.34%	82.42%	80.60%
8	81.87%	81.58%	81.87%	81.72%
9	78.02%	78.02%	78.02%	78.02%
10	54.40%	77.46%	54.40%	61.20%
Rata - Rata	76.83%	78.98%	76.83%	77.39%

Pada tabel 3. menunjukkan hasil akurasi model *Decision Tree* setelah penyeimbangan data dengan metode *ADASYN* dari 10 *fold* evaluasi. Secara umum, kinerja model meningkat dengan akurasi berkisar antara 75,28% hingga 81,58%, dan *precision* mencapai 78,98%. *Recall* juga konsisten tinggi, mencerminkan kemampuan model mengenali kelas positif dengan baik. *F1-score* menunjukkan performa yang stabil dengan rata-rata di atas 75%, kecuali pada *fold* ke-10 yang lebih rendah (61,20%). Hasil ini menunjukkan bahwa *ADASYN* berhasil meningkatkan stabilitas dan keandalan model terhadap data tidak seimbang. Evaluasi model secara keseluruhan cukup baik dengan skor di atas 76,83% untuk semua metrik, mengindikasikan bahwa fitur yang digunakan sudah efektif untuk klasifikasi. Meskipun tanpa teknik *Forward Selection*, perhitungan manual *Confusion Matrix* dilakukan untuk mengonfirmasi keakuratan hasil evaluasi.

Tabel 4. *Confusion Matrix*

	<i>Predicted Positive (0)</i>	<i>Predicted Negative (1)</i>
<i>Actual Positive (0)</i>	1329	237
<i>Actual Negative (1)</i>	185	71



Gambar 2. *Confusion Matrix Decision Tree*

Pada tabel 4. dan Gambar 2. di atas menunjukkan performa model *Decision Tree* dalam mengklasifikasikan tingkat keparahan kecelakaan lalu lintas, dengan dua kelas utama: kelas 0 (kategori Ringan dan Sedang) dan kelas 1 (kategori Berat). Dari total prediksi, sebanyak 1329 data kelas 0 berhasil diprediksi dengan benar (*true negative*), sedangkan 237 data dari kelas yang sama salah diprediksi sebagai kelas 1 (*false positive*). Untuk kelas 1, model berhasil mengklasifikasikan 71 data secara tepat (*true positive*), namun terdapat 185 data yang salah diprediksi sebagai kelas 0 (*false negative*). Hasil ini menunjukkan bahwa model memiliki akurasi tinggi dalam mengenali kelas mayoritas, tetapi masih mengalami kesulitan dalam mengidentifikasi kelas minoritas, yang tercermin dari nilai *false negative* yang cukup tinggi. Oleh karena itu, diperlukan pendekatan tambahan seperti penyeimbangan data agar model lebih sensitif terhadap kategori kecelakaan berat.

$$Accuracy = \frac{1329 + 71}{1329 + 71 + 237 + 185} = 0.7683 = 76.83\%$$

3.4 Hasil Permodelan Kombinasi dengan *Forward Selection* dan *ADASYN*

Fitur-fitur terbaik dipilih menggunakan metode *Forward Selection*. Dengan cara ini, model hanya dilatih menggunakan atribut-atribut paling informatif. Hasil pengujian menunjukkan bahwa penerapan *Forward Selection* mampu meningkatkan akurasi hingga rata-rata 99,06%, jauh lebih baik dibanding model awal. Selain akurasi, nilai *precision*, *recall*, dan *f1-score* juga meningkat secara signifikan, memperkuat bahwa seleksi fitur berpengaruh positif terhadap kinerja model.

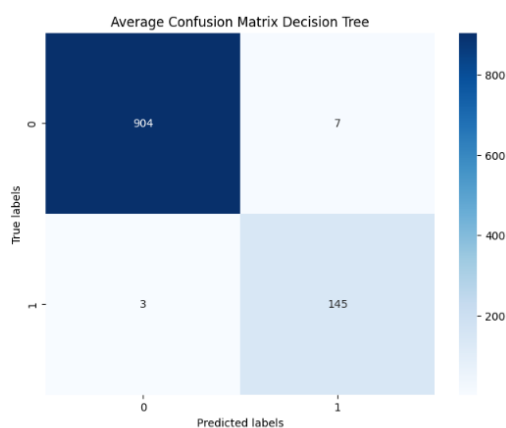
Tabel 5. Hasil Pengujian Model *Decision Tree* Dengan *Forward Selection*

Fold	Akurasi	Precision	Recall	F1 - Score
1	98.11%	98.11%	98.11%	98.11%
2	100%	100%	100%	100%
3	98.11%	98.34%	98.11%	98.16%
4	98.11%	98.11%	98.11%	98.11%
5	99.06%	99.12%	99.06%	99.07%
6	99.06%	99.12%	99.06%	99.07%
7	100%	100%	100%	100%
8	100%	100%	100%	100%
9	99.06%	99.07%	99.06%	99.04%
10	99.05%	99.11%	99.05%	99.06%
Rata - rata	99.06%	99.10%	99.06%	99.06%

Hasil pada Tabel 5. menunjukkan bahwa model *Decision Tree* dengan seleksi fitur *Forward Selection* pada *10K-fold* berhasil mencapai akurasi tinggi, dengan rata-rata 99,06%. Tingkat akurasi ini mencerminkan kemampuan model dalam mengklasifikasikan data kecelakaan lalu lintas di Kota Samarinda dengan tingkat kesalahan yang sangat rendah. Penerapan *Forward Selection* membantu model mengidentifikasi atribut paling berpengaruh, sehingga meningkatkan akurasi prediksi. Evaluasi menunjukkan adanya peningkatan kinerja setelah seleksi fitur, dan untuk memastikan keakuratan hasil, dilakukan perhitungan manual *Confusion Matrix*.

Tabel 6. Confusion Matrix

	Predicted Positive (0)	Predicted Negative (1)
Actual Positive (0)	904	7
Actual Negative (1)	3	145



Gambar 3. Confusion Matrix Decision Tree

Pada tabel 6. dan gambar 3. di atas menunjukkan performa model *Decision Tree* setelah penerapan *ADASYN* dan *Forward Selection*. Model mampu mengklasifikasikan data dengan sangat baik,

dengan 904 data kelas 0 (Ringan dan Sedang) berhasil diprediksi dengan benar dan hanya 7 yang salah klasifikasi. Sementara itu, dari kelas 1 (Berat), sebanyak 145 data berhasil diklasifikasikan dengan benar dan hanya 3 yang salah. Hasil ini mencerminkan tingkat akurasi dan presisi yang sangat tinggi, serta menunjukkan bahwa kombinasi metode *ADASYN* dan *Forward Selection* secara efektif meningkatkan kinerja klasifikasi model.

$$Accuracy = \frac{145 + 904}{145 + 904 + 7 + 3} = 0.9906 = 99.06\%$$

Meskipun akurasi mencapai 99,06% setelah penerapan *Forward Selection* dan *ADASYN*, hal ini berpotensi menunjukkan *overfitting*, di mana model terlalu menyesuaikan diri dengan data latih dan kurang mampu menggeneralisasi data baru. Evaluasi lanjutan dengan data uji independen diperlukan untuk memastikan validitas model.

3.2 Evaluasi dan Perbandingan Hasil

Dibandingkan dengan model tanpa *ADASYN* dan *Forward Selection* (69,44%). Nilai penerapan metode *ADASYN* meningkatkan akurasi menjadi 76,83%. Saat ditambahkan dengan metode *Forward Selection* setelah metode penyeimbangan data, akurasi bertambah drastis ke 99,06%.

Tabel 7. Perbandingan Hasil Akurasi Pengujian *Decision Tree*

Fold	Decision Tree (Tanpa <i>Forward Selection</i> dan <i>ADASYN</i>)	Decision Tree setelah <i>ADASYN</i> (Sebelum <i>Forward Selection</i>)	Decision Tree (Sesudah <i>Forward Selection</i> dan <i>ADASYN</i>)	Perbandingan Hasil DT + DT Setelah <i>ADASYN</i>	Perbandingan Hasil DT Setelah <i>ADASYN</i> + DT Setelah FS & <i>ADASYN</i>
1	67.21%	79.23%	98.11%	+12.02%	+18.88%
2	50.82%	81.42%	100%	+30.60%	+18.58%
3	73.63%	73.63%	98.11%	+0.00%	+24.48%
4	80.77%	80.77%	98.11%	+0.00%	+17.34%
5	80.22%	80.22%	99.06%	+0.00%	+18.84%
6	76.37%	76.37%	99.06%	+0.00%	+22.69%
7	86.81%	82.42%	100%	-4.39%	+17.58%
8	81.32%	81.87%	100%	+0.55%	+18.13%
9	48.90%	78.02%	99.06%	+29.12%	+21.04%
10	48.35%	54.40%	99.05%	+6.05%	+44.65%
Rata - rata	69.44%	76.83%	99.06%	+10,64%	+28,93%

Penerapan metode *ADASYN* dan *Forward Selection* pada tabel 7. secara bertahap memberikan peningkatan akurasi pada model *Decision Tree* di setiap *fold*. *Confusion Matrix* juga menunjukkan bahwa pada model mampu mengklasifikasikan kelas kecelakaan berat dengan lebih akurat setelah penerapan tersebut.

3.9 Pembahasan

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa poin utama yang menunjukkan efektivitas metode yang digunakan dalam meningkatkan kinerja klasifikasi data kecelakaan lalu lintas. Adapun poin – poin penting yang dapat disimpulkan dari penerapan metode *Forward Selection* dan *ADASYN* dengan algoritma *Decision Tree* adalah sebagai berikut:

- a) Berdasarkan hasil penelitian ini, fitur yang paling berpengaruh dalam menentukan tingkat keparahan kecelakaan lalu lintas di Kota Samarinda adalah jumlah korban meninggal dunia, jumlah luka berat, jumlah luka ringan, dan cuaca, sebagaimana diidentifikasi melalui metode *Forward Selection* dengan algoritma *Decision Tree*. Fitur-fitur tersebut terbukti mampu meningkatkan akurasi model klasifikasi secara signifikan. Temuan ini sejalan dengan penelitian oleh [19] di Kota Pekanbaru yang menunjukkan bahwa atribut waktu kejadian, usia korban, jenis kendaraan yang terlibat, jenis kecelakaan, dan jenis lawan tabrak memberikan kontribusi signifikan dalam mengklasifikasikan kondisi korban. Dengan demikian, baik kondisi korban maupun faktor lingkungan terbukti menjadi indikator utama dalam menentukan tingkat keparahan kecelakaan lalu lintas.
- b) Hasil penelitian ini menunjukkan bahwa penerapan tiga metode sekaligus, yaitu *Decision Tree*, *ADASYN* dan *Forward Selection*, terbukti sangat efektif meningkatkan kinerja pada model. Awalnya *Decision Tree* tanpa penanganan ketidakseimbangan dan tanpa seleksi fitur hanya menghasilkan akurasi 69,44%, *precision* 79,01%, *recall* 69,44% dan *f1-score* 71,79%. Setelah menerapkan metode *ADASYN*, metrik evaluasi meningkat signifikan menjadi akurasi 76,83%, *precision* 78,98%, *recall* 76,83% dan *f1-score* 77,39%. Peningkatan lebih besar tercapai setelah menambahkan seleksi fitur *Forward Selection*, menghasilkan hasil akhir akurasi 99,06%, *precision* 99,10%, *recall* 99,06% dan *f1-score* 99,06%. Penelitian lain oleh [9] pada klasifikasi jenis kecelakaan lalu lintas menggunakan *Decision Tree* memperoleh hasil akurasi sebesar 93%, Sementara evaluasi metrik lainnya juga menunjukkan hasil tinggi. Namun penelitian tersebut tidak

memanfaatkan teknik *balancing* atau seleksi fitur. Berbeda dengan studi-studi sebelumnya yang umumnya menerapkan salah satu metode saja, penelitian ini secara orisinal menggabungkan dua pendekatan sekaligus *ADASYN* untuk menyeimbangkan distribusi kelas dan *Forward Selection* untuk memilih fitur paling relevan ke dalam model *Decision Tree* dalam konteks klasifikasi kecelakaan lalu lintas. Hingga saat ini, belum ditemukan studi yang secara eksplisit menguji kombinasi kedua metode ini secara bersamaan, terutama pada kasus di Indonesia. Hal ini menjadi kontribusi ilmiah baru yang ditawarkan dalam penelitian ini. Sebagai perbandingan, penelitian [20] menunjukkan bahwa penerapan metode *Forward Selection* mampu meningkatkan performa klasifikasi pada kasus penyakit gagal jantung. Sementara itu, penelitian [21] membuktikan bahwa teknik *ADASYN* dapat meningkatkan performa model klasifikasi penyakit diabetes secara signifikan. Hasil-hasil ini mendukung bahwa masing-masing metode memiliki potensi besar, namun belum ada yang secara terpadu mengaplikasikan keduanya dalam domain kecelakaan lalu lintas. Oleh karena itu, kombinasi *ADASYN* dan *Forward Selection* dalam penelitian ini merupakan pendekatan yang baru sekaligus efektif, terutama dalam mendeteksi kasus kecelakaan berat yang sebelumnya sulit teridentifikasi akibat ketidakseimbangan data.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan maka dapat ditarik beberapa kesimpulan sebagai berikut:

- a) Hasil dari penerapan seleksi fitur dengan menggunakan *Forward Selection* pada dataset kecelakaan lalu lintas Kota Samarinda didapat 4 fitur terpilih yang memiliki pengaruh paling signifikan terhadap klasifikasi kecelakaan lalu lintas yaitu jumlah meninggal dunia, jumlah luka berat, jumlah luka ringan dan cuaca.
- b) Kombinasi metode *Forward Selection* dan *ADASYN* terbukti efektif dalam meningkatkan akurasi klasifikasi data kecelakaan lalu lintas dengan algoritma *Decision Tree*. Penerapan *ADASYN* terlebih dahulu mampu menyeimbangkan distribusi kelas dalam dataset, sehingga model

menjadi lebih adil dalam mempelajari kelas minoritas. Selanjutnya, penggunaan *Forward Selection* memungkinkan model untuk fokus pada fitur-fitur yang paling relevan, mengurangi kompleksitas model, dan meningkatkan efisiensi pemrosesan data. Hasil akhir menunjukkan peningkatan akurasi yang signifikan, dari 69,44% pada model awal menjadi 99,06% setelah penerapan kedua metode tersebut. Hal ini menegaskan bahwa pendekatan bertahap yang menggabungkan penyeimbangan data dan seleksi fitur dapat secara substansial meningkatkan performa model klasifikasi.

5. SARAN

Berdasarkan hasil dan pembahasan yang telah diperoleh, terdapat beberapa saran yang diharapkan dapat menjadi masukan untuk pengembangan penelitian selanjutnya maupun untuk penerapan hasil penelitian ini secara lebih optimal, sebagai berikut:

1. Penelitian selanjutnya di sarankan untuk membandingkan kinerja algoritma *Decision Tree* dengan algoritma yang lain seperti *Neural Networks*, *XGBoost*, atau *Gradient Boosting*. Hal ini Bertujuan untuk mengetahui apakah algoritma lain dapat memberikan hasil klasifikasi yang lebih akurat dan stabil terhadap data kecelakaan lalu lintas.
2. Selain metode *ADASYN*, penelitian lanjutan dapat mencoba teknik balancing lainnya seperti *SMOTE*, *Tomek Links*, atau kombinasi *oversampling* atau *undersampling* untuk melihat dampaknya terhadap performa model. Penambahan fitur baru seperti waktu kejadian (jam sibuk dan jam tidak sibuk), jenis kendaraan, atau juga analisis spasial berdasarkan lokasi kecelakaan juga dapat meningkatkan ketepatan prediksi dan memberikan wawasan yang lebih mendalam.

DAFTAR PUSTAKA

- [1] BPS Prov. Kaltim, "Jumlah Kecelakaan Lalu Lintas 2021-2023," Badan Pusat Statistik Provinsi Kalimantan Timur. Accessed: Mar. 21, 2025.
<https://kaltim.bps.go.id/id/statistics-table/2/NDE2IzI%3D/jumlah-kecelakaan-lalu-lintas-.html>
- [2] Iswanto and T. Romadhani, "Polresta Samarinda Ungkap Data Laka Lantas 2023, Kelalaian Jadi Faktor Utama Penyebab Terjadinya Kecelakaan," *nomorsatukaltim*. Accessed: Mar. 21, 2025.
<https://nomorsatukaltim.disway.id/read/38918/polresta-samarinda-ungkap-data-laka-lantas-2023-kelalaian-jadi-faktor-utama-penyebab-terjadinya-kecelakaan>
- [3] Adit Mustafa, "233 Kasus Laka Lantas Tercatat di Samarinda Sepanjang 2024 dan 28 Persen Meninggal Dunia," *insitekalim*. Accessed: Mar. 21, 2025.
<https://insitekalim.com/233-kasus-laka-lantas-tercatat-di-samarinda-sepanjang-2024-dan-28-persen-meninggal-dunia/>
- [4] A. H. Sheng Dong, Afaq Khattak, Irfan Ullah, Jibiao Zhou, "Predicting and Analyzing Road Traffic Injury Severity Using Boosting-Based Ensemble Learning Models with SHAPley Additive exPlanations," *Int. J. Environ. Res. Public Health*, vol. 19, no. 5, p. 23, 2022, doi: 10.3390/ijerph19052925.
- [5] C. Nauri, D. M. Hastamo, D. A. Saputra, and V. Yovian, "Implementation of Machine Learning on Android Based Application for Classifying the Types of Injuries on Traffic Accident Implementasi Machine Learning Menggunakan Aplikasi Android untuk Klasifikasi Jenis Luka Cedera pada Kecelakaan Lalu Lintas," *RADIANT J. Applied, Soc. Educ. Stud.*, vol. 6, no. 1, pp. 1–8, 2025.
- [6] B. L. Fauzan, T. Agustin, and A. M. H. Mahmudah, "Prediksi Klasifikasi Kecelakaan Lalu Lintas di Kota Surakarta dengan Menggunakan Metode Regresi Logistik Multinomial," *Sustain. Civ. Build. Manag. Eng. J.*, vol. 1, no. 4, pp. 1–9, 2024, doi: 10.47134/scbmej.v1i4.3159.
- [7] H. Triana and U. Enri, "Penerapan Deep Learning Pada Kamera Pengawas Jalan Raya Dalam Mendeteksi Kecelakaan," *Syntax J. Inform.*, vol. 11, no. 02, pp. 46–55, 2022, doi: 10.35706/syji.v11i02.6356.
- [8] R. N. Ramadhon, A. Ogi, A. P. Agung, R. Putra, S. S. Febrihartina, and U. Firdaus, "Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak

- Aktif pada Data Bank,” *Karimah Tauhid*, vol. 3, pp. 1860–1874, 2024, doi: <https://doi.org/10.30997/karimahtauhid.v3i2.11952>.
- [9] K. Octavia, F. Yulianto, and T. A. Y. Siswa, “Penentuan jenis kecelakaan lalu lintas di kota samarinda menggunakan metode,” *Didakt. J. Ilm. PGSD FKIP Univ. Mandiri*, vol. 11, no. 1, pp. 212–226, 2025, doi: <https://doi.org/10.36989/didaktik.v11i01.5589>.
- [10] H. Hafizan and A. N. Putri, “Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi Balita Di Kabupaten Simalungun,” *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 1, no. 2, pp. 68–72, 2020, doi: [10.30645/kesatria.v1i2.23](https://doi.org/10.30645/kesatria.v1i2.23).
- [11] L. R. A. Tarigan and Dahlan, “OPTIMALISASI FITUR DENGAN FORWARD SELECTION PADA ESTIMASI TINGKAT PENYAKIT PARU-PARU MENGGUNAKAN ALGORITMA,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 5, pp. 10341–10348, 2024, doi: <https://doi.org/10.36040/jati.v8i5.11064>.
- [12] I. Ubaedi and Y. M. Djaksana, “Optimasi Algoritma C4.5 Menggunakan Metode Forward Selection Dan Stratified Sampling Untuk Prediksi Kelayakan Kredit,” *JSiI (Jurnal Sist. Informasi)*, vol. 9, no. 1, pp. 17–26, 2022, doi: [10.30656/jsii.v9i1.3505](https://doi.org/10.30656/jsii.v9i1.3505).
- [13] A. B. Alpriansah and Y. Ramdhani, “Optimasi Fitur dengan Forward Selection pada Estimasi Tingkat Obesitas menggunakan Random Forest,” *Sist. J. Sist. Inf.*, vol. 12, no. 3, pp. 860–873, Sep. 2023, doi: [10.32520/stmsi.v12i3.3125](https://doi.org/10.32520/stmsi.v12i3.3125).
- [14] I. R. Lina and S. Retnowati, “Metode K-Nearest Neighbor Berbasis Forward Selection untuk Identifikasi Status Stunting pada Balita,” *J. Comput. Sci. Inf. Technol.*, vol. 5, no. 3, pp. 695–704, 2024, doi: <https://doi.org/10.37859/coscitech.v5i3.8061>.
- [15] A. M. A. Rahim, B. P. Hartato, A. Ridwan, and F. Asharudin, “Machine Learning-Based Approach for HIV / AIDS Prediction: Feature Selection and Data Balancing Strategy,” *J. Appl. Informatics Comput.*, vol. 9, no. 2, pp. 338–347, 2025, doi: <https://doi.org/10.30871/jaic.v9i2.9125>.
- [16] M. T. T. Br Sirait, N. S. Fathonah, and M. N. Fauzan, “PEMANFAATAN ALGORITMA ADASYN DAN SUPPORT VECTOR MACHINE DALAM MENINGKATKAN AKURASI PREDIKSI KANKER PARU-PARU,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 5, pp. 8773–8778, 2024, doi: <https://doi.org/10.36040/jati.v8i5.10752>.
- [17] W. Hidayat, M. Ardiansyah, and A. Setyanto, “Pengaruh Algoritma ADASYN dan SMOTE terhadap Performa Support Vector Machine pada Ketidakseimbangan Dataset Airbnb,” *Edumatic J. Pendidik. Inform.*, vol. 5, no. 1, pp. 11–20, 2021, doi: [10.29408/edumatic.v5i1.3125](https://doi.org/10.29408/edumatic.v5i1.3125).
- [18] J. Al Amien, Yoze Rizki, and Mukhlis Ali Rahman Nasution, “Implementasi Adasyn Untuk Imbalance Data Pada Dataset UNSW-NB15 Adasyn Implementation For Data Imbalance on UNSW-NB15 Dataset,” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 3, pp. 242–248, 2022, doi: [10.37859/coscitech.v3i3.4339](https://doi.org/10.37859/coscitech.v3i3.4339).
- [19] S. N. Abdullah, S. Kurnia Gusti, F. Wulandari, and F. Syafria, “Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas Menggunakan Naïve Bayes,” *Progresif J. Ilm. Komput.*, vol. 19, no. 2, p. 858, 2023, doi: [10.35889/progresif.v19i2.1348](https://doi.org/10.35889/progresif.v19i2.1348).
- [20] K. Kusnawi and Khrisna Irham Fadhil Pratama, “Komparasi Algoritma Supervised Learning dan Feature Selection pada Klasifikasi Penyakit Gagal Jantung,” *Indones. J. Comput. Sci.*, vol. 12, no. 6, pp. 3722–3733, Dec. 2023, doi: [10.33022/ijcs.v12i6.3487](https://doi.org/10.33022/ijcs.v12i6.3487).
- [21] M. I. Anugrah, J. Zeniarja, and D. Setiawan, “Peningkatan Performa Model Hard Voting Classifier dengan Teknik Oversampling ADASYN pada Penyakit Diabetes,” *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, pp. 290–299, 2024, doi: [10.29408/edumatic.v8i1.25838](https://doi.org/10.29408/edumatic.v8i1.25838).