JURNAL BUFFER INFORMATIKA
Volume 11 Nomor 2, Oktober 2025
p-ISSN : 2527-4856, e-ISSN : 2614-5413
https://journal.fkom.uniku.ac.id/index.php/buffer

Penerapan Metode Forward Selection dan ADASYN pada Algoritma SVM untuk Klasifikasi Data Kecelakaan Lalu Lintas

Muhammad Fadly Ramadhani[1)] , Taghfirul Azhima Yoga Siswa[2)], Wawan Joko Pranoto[3)]

[1,2,3),] Fakultas Saint dan Teknologi, Universitas Muhammadiyah Kalimantan Timur,
Jl. Ir. H. Juanda No.15, Sidodadi, Kec. Samarinda Ulu, Kota Samarinda, Kalimantan Timur 75124
Email :2111102441117@umkt.ac.id[1)],tay758@umkt.ac.id[2)],wjp337@umkt.ac.id[3)]

Abstrak

Di Indonesia, khususnya kota Samarinda, angka kecelakaan lalu lintas menunjukkan tren peningkatan yang mengkhawatirkan. Jumlah kecelakaan meningkat dari 97 kasus pada tahun 2021 menjadi 102 kasus pada tahun 2022, dan mencapai 173 kasus pada tahun 2023. Penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi tingkat kecelakaan lalu lintas dengan mengimplementasikan algoritma *Support Vector Machine (SVM)* yang dikombinasikan dengan metode seleksi fitur *Forward Selection* dan teknik *oversampling ADASYN*. Dataset yang digunakan merupakan data kecelakaan dari Polresta Samarinda periode 2020–2024 dengan 35 atribut, yang diseleksi menjadi 13 atribut relevan. Penelitian dilakukan melalui tahapan data *pre-processing*, balancing, pemodelan *SVM*, serta evaluasi menggunakan *10-fold cross-validation*. Hasil pengujian menunjukkan bahwa penerapan *ADASYN* mampu meningkatkan akurasi model dari 70,75% menjadi 96,38%. Peningkatan lebih lanjut dicapai dengan *Forward Selection*, menghasilkan akurasi hingga 98,00%. Temuan ini membuktikan bahwa seleksi fitur dan penyeimbangan kelas memiliki kontribusi signifikan dalam memperkuat performa model klasifikasi *SVM* untuk analisis kecelakaan lalu lintas.

Kata Kunci : SVM; ADASYN; Forward Selection; Kecelakaan lalu lintas; Klasifikasi

Abstract

In Indonesia, particularly in the city of Samarinda, the number of traffic accidents has shown a worrying upward Indonesia. The number of accidents increased from 97 cases in 2021 to 102 cases in 2022, and reached 173 cases in 2023. This study aims to improve the accuracy of traffic accident severity classification by implementing the Support Vector Machine (SVM) algorithm combined with the Forward Selection feature selection method and the ADASYN oversampling technique. The dataset used consists of accident data from the Samarinda Police Department for the 2020–2024 period, with 35 attributes that were reduced to 13 relevant features. The research was conducted through stages of data pre-processing, balancing, SVM modeling, and evaluation using 10-fold cross-validation. The test results show that the application of ADASYN increased the model's accuracy from 70.75% to 96.38%. Further improvement was achieved with Forward Selection, resulting in an accuracy of up to 98.00%. These findings demonstrate that feature selection and class balancing significantly contribute to enhancing the performance of the SVM classification model for traffic accident analysis.

Keywords: SVM; ADASYN; Forward Selection; Traffic Accidents; Classification

1. PENDAHULUAN

Di Indonesia, khususnya Kota Samarinda, angka kecelakaan lalu lintas menunjukkan tren peningkatan yang mengkhawatirkan. Jumlah kecelakaan meningkat dari 97 kasus pada tahun 2021 menjadi 102 kasus pada tahun 2022, dan mencapai 173 kasus pada tahun 2023 [1] Selain itu, data dari Insitekaltim mencatat bahwa pada tahun 2024 terdapat 233 kasus kecelakaan, dengan 28% di antaranya berujung pada korban meninggal dunia [2] Faktor penyebab utama kecelakaan lalu lintas tersebut meliputi, berkendara dengan kecepatan tinggi, tidak mematuhi rambu lalu lintas, penggunaan ponsel sambil berkendara dan berkendara dalam keadaan mabuk [3].

Penerapan *Machine Learning* menjadi salah satu pendekatan yang efektif dalam menganalisis

kecelakaan lalu lintas. Salah satu metode yang bisa digunakan adalah klasifikasi, dibuktikan dengan beberapa penelitian sebelumnya seperti penerapan algoritma *KNN*, dan *Naïve Bayes* untuk klasifikasi data kecelakaan lalu lintas, dengan hasil akurasi 75,86%, dan 79,31% (Nabila Abda Salsabila et al., 2023). Algoritma klasifikasi ini juga telah terbukti efektif dalam menangani kasus medis lainnya, seperti penyakit Kanker Payudara dengan akurasi *Naïve bayes* (72%), dan *SVM* (88%) [6], [7].

Algoritma *Support Vector Machine (SVM)* dikenal sebagai salah satu metode klasifikasi yang memiliki hasil tinggi dalam melakukan prediksi pengklasifikasian [8]. Algoritma *Support Vector Machine (SVM)* juga efektif digunakan untuk mengidentifikasi berbagai faktor yang memengaruhi tingkat keparahan kecelakaan [9]. Hasil pengujian menunjukkan bahwa *SVM* mampu mencapai akurasi sebesar 94,62% dalam proses klasifikasi. Selain itu, penelitian terkait kecelakaan lalu lintas di Lombok Timur melaporkan bahwa algoritma ini berhasil memperoleh akurasi sebesar 93,66% [10], [11]. Meskipun persentase yang diperoleh cukup tinggi, namun sebagian besar proses pemrosesannya masih menggunakan dataset dengan dimensi rendah.

Dataset berdimensi tinggi adalah dataset yang memiliki banyak fitur. Dalam penerapannya, algoritma seringkali mengalami penurunan performa yang signifikan ketika dihadapkan dengan dataset berdimensi tinggi. Sepert penelitian oleh [12]. hasil menunjukkan bahwa dalam mengklasifikasikan kelayakan air minum, algoritma KNN dan SVM belum memberikan performa yang optimal, dengan tingkat akurasi masing-masing sebesar 65% dan 69%. Untuk mengatasi masalah jumlah dimensi pada dataset, diperlukan seleksi fitur yang bertujuan untuk mengurangi jumlah fitur dengan hanya mempertahankan fitur-fitur yang memiliki relevansi paling tinggi dibandingkan dengan fitur lainnya.

Metode *Forward Selection* merupakan teknik seleksi fitur yang dapat digunakan untuk memilih atribut yang paling berpengaruh dalam klasifikasi. Penerapan seleksi fitur *forward selection* terbukti berpengaruh terhadap peningkatan akurasi klasifikasi yang signifikan, peningkatan klasifikasi sebesar 94,84% dengan membuang beberapa fitur yang tidak relevan terhadap klasifikasi dan menjadikannya hasil yang lebih baik (Fanani, 2020).

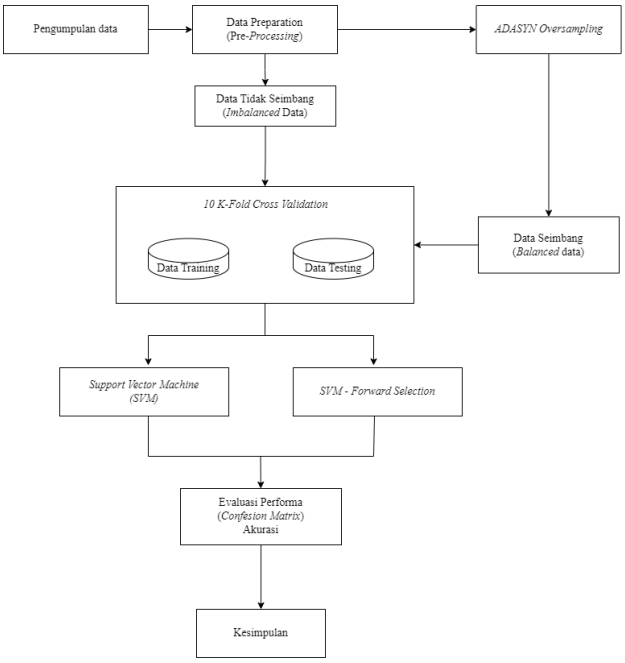
Salah satu tantangan utama lainnya dalam analisis data kecelakaan adalah ketidakseimbangan kelas (*class imbalanc*e), di mana jumlah kasus kecelakaan berat jauh lebih sedikit dibandingkan kecelakaan ringan atau tanpa korban. Ketidakseimbangan ini dapat menyebabkan bias dalam prediksi, di mana algoritma cenderung mengabaikan kelas minoritas yang justru penting untuk mitigasi risiko kecelakaan [14]. Untuk mengatasi masalah ini, metode *Adaptive Synthetic Sampling (ADASYN)* terbukti efektif dalam menangani ketidakseimbangan kelas dengan membuat sampel sintetis yang fokus pada batas keputusan antar kelas. *ADASYN* berhasil meningkatkan akurasi, spesifisitas, dan sensitivitas masing-masing 72,88%, 73,42%, dan 71,79% dalam klasifikasi diabetes [15].

Berdasarkan tinjauan literatur yang telah dilakukan, penelitian ini mengusulkan penerapan kombinasi algoritma *SVM* dengan metode seleksi fitur *Forward Selection* dan teknik oversampling *ADASYN* untuk meningkatkan kinerja *SVM* dalam klasifikasi kecelakaan lalu lintas di Samarinda, terutama pada dataset dengan dimensi tinggi dan ketidakseimbangan data. Oleh karena itu, penelitian ini dianggap inovatif dan diharapkan dapat memberikan kontribusi dalam peningkatan akurasi klasifikasi meskipun dihadapkan dengan tantangan dataset berdimensi tinggi dan ketidakseimbangan data pada kecelakaan lalu lintas.

2. METODE PENELITIAN

2.1 Prosedur Penelitian

Berikut adalah prosedur penelitian yang akan diterapkan dalam studi ini:



Gambar 1. 1 Alur Penelitian

2.2 Pengumpulan Data

Pada penelitian ini menggunakan data Laka Lantas yang bersifat privasi yang didapatkan secara langsung di Polresta Kota Samarinda. Polresta Kota Samarinda dengan izin akses yang sudah disetujui. Pada periode tahun 2020-2024. Dari instansi tersebut mendapatkan fitur-fitur yang di gunakan meliputi nomor laka, polres, tanggal kejadian, waktu kejadian, tanggal dilaporkan, waktu dilaporkan, tingkat kecelakaan, jumlah meninggal dunia, jumlah luka berat, jumlah luka ringan, koordinat GPS-Lintang, koordinat GPS-Bujur, titik acuan/referensi, jarak ke lokasi, 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. Lokasi penelitian di Polresta kota Samarinda Jl. Slamet Riyadi No.1, Karang Asam Ulu, Kec. Sungai Kunjang, kota Samarinda, Kalimantan Timur 75126.

2.3 Data *Pre-Processing*

Proses pengolahan data kecelakaan lalu lintas yang diperoleh dari Polresta kota Samarinda harus melalui tahap pengolahan lebih lanjut sebelum dilakukan pemodelan. Tujuannya adalah untuk menghilangkan bagian data yang tidak relevan dan memastikan kualitas hasil dalam proses analisis data. Tahapan pengolahan data ini sangat krusial untuk menjamin keakuratan dalam proses *Machine Learning*. Pada tahapan ini ada beberapa langkah yaitu data *Selection*, data *Cleaning,* data *Transformation* dan data *balancing.*

a) *Data selection* adalah proses memilih subset data relevan dari data mentah untuk dianalisis, dengan tujuan meningkatkan efisiensi, mengurangi kompleksitas, dan memastikan kualitas hasil. Pemilihan didasarkan pada atribut, kondisi tertentu, atau kualitas data.

b) *Data cleaning* adalah proses mendeteksi dan memperbaiki kesalahan/inkonsistensi dalam data, seperti data hilang, duplikat, format salah, outlier, atau nilai tidak logis, guna meningkatkan keakuratan dan keandalan analisis.

c) *Transformasi data* dilakukan dengan LabelEncoder, yaitu metode mengubah nilai kategorikal menjadi numerik agar algoritma machine learning dapat memproses data secara optimal [16].

d) *Data balancing* adalah metode untuk mengatasi ketidakseimbangan kelas dalam dataset, yang bisa mengganggu kinerja model karena dominasi kelas tertentu.

2.4 Pembagian Data *Training* dan Data *Testing*

Dataset dibagi menjadi data pelatihan dan pengujian. Data pelatihan digunakan untuk melatih model, sedangkan data pengujian untuk mengukur kinerjanya. Evaluasi dilakukan menggunakan teknik 10-Fold Cross-Validation, yang membagi data menjadi 10 bagian dan menguji model secara bergantian untuk memperoleh hasil evaluasi yang lebih akurat [16].

Tabel 2.1 10-Fold Cross-Validation

1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10

Pada Tabel 2.1 skema 10 k-fold cross validation ditunjukkan, di mana setiap baris mewakili satu iterasi. Data yang ditandai dengan warna merah menunjukkan data yang digunakan sebagai data uji pada iterasi tersebut, sedangkan blok berwarna putih menunjukkan data yang digunakan untuk pelatihan. Proses ini diulang sebanyak 10 kali, dan hasil evaluasi dari setiap fold dihitung rata-ratanya untuk memberikan estimasi kinerja model yang lebih akurat.

2.5 Permodelan *Support Vector Machine (SVM)*

Support Vector Machine (SVM) adalah algoritma machine learning untuk klasifikasi dan regresi yang mencari hyperplane optimal sebagai pemisah antar kelas. SVM menggunakan hyperplane linear atau non-linear tergantung pada pola data [17]Algoritma ini unggul karena mampu bekerja baik pada data kecil maupun besar, serta efektif pada data dengan banyak atribut dan mudah diimplementasikan [18]

$$f(x) = wx + b = 0$$

Keterangan :
w :Menentukan orientasi *hiperplane*.
x :Merupakan data masukan yang ingin di masukan
b :Menggeser *hyperplane* tanpa mengubah arahannya
f(x) :Hasil dari persamaan ini digunakan untuk menentukan posisi *hyperplane*

2.6 Permodelan *SVM* dengan *Forward Selection*

Penerapan *Support Vector Machine (SVM)* dengan metode *Forward Selection* telah terbukti efektif dalam meningkatkan kinerja prediksi di berbagai bidang *pattern recognition* [19].

$$R^2_{adj} = 1 - \left(\frac{(1-R^2)(n-1)}{n-p-1}\right)$$

Keterangan :
R^2 : Kofisien determinasi (menguji sejauh mana variabel independent dapat menjelaskan variasi dalam variabel dependen).
n : Jumlah sampel.
p : Jumlah fitur (variabel independent) dalam model.

2.7 Evaluasi Model

Pada tahapan ini evaluasi yang diterapkan umumnya di lakukan menggunakan *confusion matrix* untuk menghitung metrik. *confusion matriks* adalah alat yang digunakan untuk menilai kinerja model klasifikasi dengan membandingkan hasil prediksi dengan nilai actual [20]

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

Keterangan :
TP *(True Positive)* : Jumlah prediksi yang benar untuk kelas positif (kelas minoritas yang benar terklasifikasikan sebagai positif).
TN *(True Negative)* : Jumlah prediksi benar untuk kelas negative (kelas mayoritas yang benar terklasifikasikan sebagai negative).
FP (*False Positive)* : Jumlah prediksi salah dimana kelas negative diklasifikasikan sebagai positif.
FN *(False Negative)* : Jumlah prediksi salah dimana kelas positif diklasifikasikan sebagai negative.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Penelitian ini menilai performa model dalam seleksi fitur serta menganalisis hasil dari penerapan *SVM, Forward Selection,* dan *ADASYN*. Data yang digunakan merupakan catatan kecelakaan lalu lintas di Kota Samarinda dengan 35 variabel prediktor dan satu target klasifikasi tingkat kecelakaan. Pendekatan ini membantu mengidentifikasi faktor dominan yang memengaruhi tingkat kecelakaan.

3.1.2 Hasil Data *Selection*

Dataset penelitian ini bersumber dari catatan kecelakaan lalu lintas di Kota Samarinda. Dari total 35 variabel prediktor, dilakukan pemilihan secara manual terhadap 13 variabel yang dinilai paling relevan dengan satu variabel target berupa tingkat keparahan kecelakaan. Variabel-variabel tersebut mencakup atribut waktu dan lokasi kejadian, kondisi lingkungan, serta penyebab kecelakaan.

Tabel 3.1 Data *Selection*

Tanggal	2018-03-12	2020-01-25	2024-12-22
Waktu	12:00	21:52	03:35
Tingkat	Ringan	Sedang	Berat
Meninggal	0	0	1
Luka Berat	0	1	0
Luka Ringan	1	0	0
GPS – Lintang	-0.462261	-0.462261	-0.486964
GPS – Bujur	117.1758	117.1758	117.1227
Cuaca	Cerah	Cerah	Hujan/grimis
Fungsi Jalan	Arteri	Arteri	Kolektor
Kondisi Jalan	Baik	Baik	Basah
Kecepatan	40 km/jam	60 km/jam	40 km/jam
Penyebab	Manusia	Manusia	Manusia

Tabel 3.1 berisi data mentah kecelakaan lalu lintas dengan atribut utama seperti waktu, tingkat keparahan, jumlah korban, lokasi, kondisi jalan dan cuaca, serta penyebab kejadian. Contohnya, kecelakaan berat pada 22 Desember 2024 pukul 03:35 terjadi saat hujan di jalan kolektor basah dan menewaskan satu orang. Data ini penting untuk analisis pola dan faktor risiko kecelakaan.

3.1.3 Hasil Data *Cleaning*

Tahap pembersihan data dimulai dengan menghapus duplikasi berdasarkan tanggal kejadian terkini. Selain itu, seluruh baris yang mengandung nilai kosong (missing values) serta atribut yang tidak

relevan dieliminasi. Dari 2.385 entri awal, tersisa sebanyak 1.167 entri bersih yang siap untuk dianalisis.

```
Jumlah missing values untuk setiap kolom:
Tanggal Kejadian             0
Waktu Kejadian               0
Tingkat Kecelakaan          17
Jumlah Meninggal Dunia       0
Jumlah Luka Berat            0
Jumlah Luka Ringan           0
Koordinat GPS - Lintang      0
Koordinat GPS - Bujur        0
Cuaca                        3
Fungsi Jalan                 5
Kondisi Permukaan Jalan      6
Batas Kecepatan Dilokasi     9
Penyebab                    84
dtype: int64
```

Gambar 3.1 Jumlah Nilai Kosong sebelum Pembersihan

Gambar 3.1 menunjukkan jumlah data yang hilang (missing values) pada setiap kolom dalam dataset kecelakaan lalu lintas. Kolom dengan data hilang terbanyak adalah "Penyebab" (84 nilai hilang), diikuti oleh "Tingkat Kecelakaan" (17), "Batas Kecepatan di Lokasi" (9), "Kondisi Permukaan Jalan" (6), "Fungsi Jalan" (5), dan "Cuaca" (3). Sementara kolom lainnya tidak memiliki nilai yang hilang. Informasi ini penting untuk proses pembersihan data sebelum analisis dilakukan.

```
Jumlah missing values untuk setiap kolom:
Tanggal Kejadian             0
Waktu Kejadian               0
Tingkat Kecelakaan           0
Jumlah Meninggal Dunia       0
Jumlah Luka Berat            0
                            ..
Cuaca                        0
Fungsi Jalan                 0
Kondisi Permukaan Jalan      0
Batas Kecepatan Dilokasi     0
Penyebab                     0
Length: 13, dtype: int64
```

Gambar 3.2 Jumlah Nilai kosong Sesudah Pembersihan

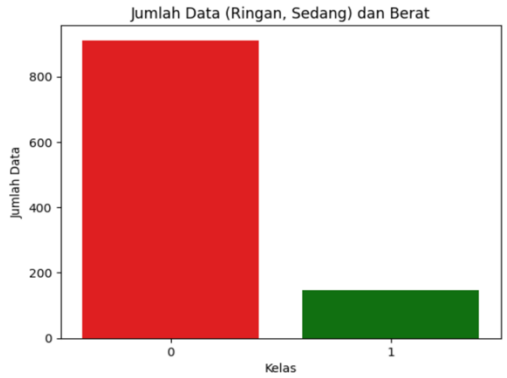
Gambar tersebut menunjukkan bahwa tidak ada missing values pada seluruh kolom dalam dataset. Artinya, semua data telah lengkap dan tidak memerlukan proses imputasi atau pengisian data yang hilang.

3.1.4 Hasil Data *Transformation*

Transformasi data dilakukan melalui encoding terhadap variabel kategorikal seperti ‘cuaca’, ‘fungsi jalan’, dan ‘penyebab’, serta konversi nilai label ‘Tingkat Kecelakaan’ menjadi dua kelas utama: kategori ringan-sedang dan kategori berat. Transformasi ini memungkinkan algoritma pembelajaran mesin memproses data secara efisien.

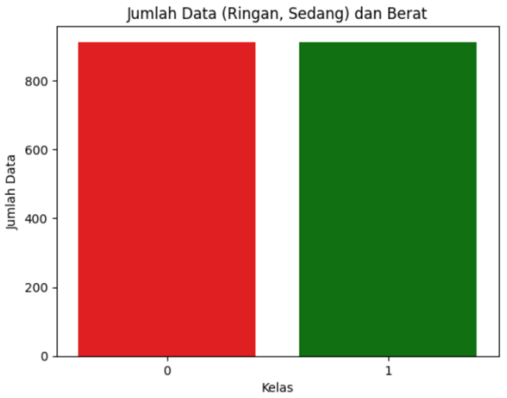
3.1.5 Hasil Data *Balancing*

Ditemukan ketidakseimbangan kelas yang cukup signifikan, di mana kelas minoritas (kategori berat) jauh lebih sedikit daripada kelas mayoritas. Untuk mengatasi hal ini, digunakan teknik *ADASYN (Adaptive Synthetic Sampling)* yang menghasilkan distribusi data yang seimbang dengan jumlah masing-masing kelas sebesar 911 dan 912 entri.



Gambar 3.3 Jumlah Kelas Sebelum *Balancing*

Pada gambar 3.3 menunjukkan ketidakseimbangan jumlah data antar kelas, di mana kategori Ringan dan Sedang (label 0) memiliki sebanyak 911 data, sementara kategori Berat (label 1) hanya terdiri dari 148 data.



Gambar 3.4 Jumlah Kelas Sesudah *Balancing*

Pada gambar 3.4 menampilkan perbandingan jumlah kelas yang telah seimbang antara kelas mayoritas dan minoritas setelah penerapan metode *ADASYN,* dengan jumlah data pada kategori Ringan dan Sedang (label 0) sebanyak 911 dan kategori Berat (label 1) sebanyak 912 data.

3.1.6 Hasil Permodelan *SVM*

Pada tahap ini, dilakukan pengujian awal menggunakan algoritma *Support Vector Machine (SVM)* dalam bahasa *Python* untuk mengevaluasi performa dasar model klasifikasi. Berdasarkan Tabel 3.6, SVM menunjukkan performa baik dan konsisten pada 10-fold cross-validation, dengan rata-rata akurasi sebesar 70,75%, yang mencerminkan kemampuan klasifikasi yang cukup andal.

Tabel 3.2 Hasil Pengujian Model *SVM*

Fold	Akurasi	*Precision*	*Recall*	*F1 - Score*
1	69.40%	21.15%	42.31%	28.21%
2	65.03%	17.24%	38.46%	23.81%
3	71.43%	22.45%	44.00%	29.73%
4	71.98%	24.00%	48.00%	32.00%
5	72.53%	24.49%	48.00%	32.43%
6	73.63%	28.30%	60.00%	38.46%
7	73.63%	29.63%	61.54%	40.00%
8	71.98%	26.42%	53.85%	35.44%
9	70.33%	20.83%	38.46%	27.03%
10	67.58%	17.65%	34.62%	23.38%

Hasil Rata – Rata	
Akurasi	70.75%
Precision	23.22%
Recall	46.92%
F1 – Score	31.05%

Model menghasilkan akurasi 70,75%, artinya sekitar 7 dari 10 prediksi sudah tepat. Namun, precision hanya 23,22%, menunjukkan banyak prediksi positif yang salah. Recall 46,92% menunjukkan model cukup baik dalam menemukan kasus positif, meski belum optimal. F1-score 31,05% mencerminkan ketidakseimbangan precision dan recall, sehingga model masih perlu ditingkatkan, misalnya melalui optimasi, seleksi fitur, atau penanganan data tidak seimbang.

3.1.7 Hasil Permodelan *SVM* setelah ADASYN

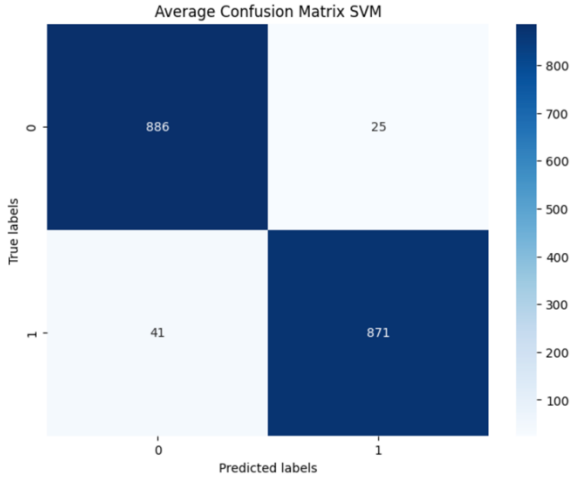
Model awal menggunakan algoritma *Support Vector Machine (SVM)* diuji dengan teknik *10-fold cross-validation.* Model menghasilkan akurasi rata-rata sebesar 96,38%, dengan precision, recall, dan F1-score yang konsisten tinggi. Hasil confusion matrix memperlihatkan kemampuan klasifikasi yang baik terhadap kedua kelas.

Tabel 3.3 Hasil Pengujian Model *SVM + ADASYN*

Fold	Akurasi	*Precision*	*Recall*	*F1 - Score*
1	97.27%	97.32%	97.27%	97.27%
2	93.44%	93.63%	93.44%	93.44%
3	97.81%	97.84%	97.81%	97.81%
4	97.80%	97.80%	97.80%	97.80%
5	95.60%	95.63%	95.60%	95.60%
6	96.70%	96.70%	96.70%	96.70%
7	95.05%	95.32%	95.05%	95.05%
8	96.15%	96.16%	96.15%	96.15%
9	97.80%	97.83%	97.80%	97.80%
10	96.15%	96.20%	96.15%	96.15%

Hasil Rata – Rata	
Akurasi	96.38%
Precision	96.44%
Recall	96.38%
F1 – Score	96.38%

Model *SVM* menunjukkan performa klasifikasi yang sangat baik dengan semua metrik di atas 96,38%, meskipun tanpa penerapan fitur seleksi. Evaluasi tambahan melalui *Confusion Matrix* memastikan keakuratan hasil tersebut.



Gambar 3.5 *Confusion Matrix*

Gambar 3.5 menunjukkan average confusion matrix dari model SVM. Hasilnya menunjukkan performa klasifikasi yang baik, dengan 886 true negative (prediksi benar untuk kelas 0) dan 871 true positive (prediksi benar untuk kelas 1). Terdapat 25 false positive dan 41 false negative, yang berarti kesalahan prediksi relatif kecil. Secara keseluruhan, model mampu membedakan kedua kelas dengan akurasi tinggi.

$$ACCURACY = \frac{886 + 871}{886 + 871 + 25 + 41} = 96\%$$

Rumus akurasi menunjukkan perhitungan akurasi model SVM, yaitu jumlah prediksi benar (886 + 871) dibagi total seluruh prediksi (886 + 871 + 25 + 41), menghasilkan akurasi sebesar 96%. Ini menunjukkan bahwa model mampu mengklasifikasikan data dengan sangat baik.

3.1.8 Hasil Permodelan SVM dengan Forward Selection

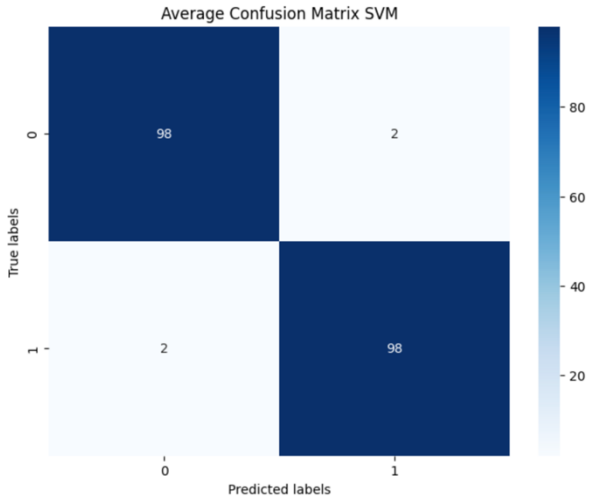
Eksperimen lanjutan mengintegrasikan metode *Forward Selection* ke dalam pemodelan *SVM* untuk mengidentifikasi fitur paling relevan. Fitur yang dipilih mencakup jumlah luka berat, jumlah luka ringan, dan kondisi cuaca. Hasil evaluasi menunjukkan peningkatan akurasi menjadi rata-rata 98,00%, dengan precision 98,18% dan F1-score 97,99%. Ini membuktikan bahwa seleksi fitur mampu meningkatkan efektivitas model dengan mengurangi kompleksitas tanpa mengorbankan akurasi.

Tabel 3.4 Hasil Pengujian Model *SVM* dengan *Forward Selection*

Fold	Akurasi	*Precision*	*Recall*	*F1 - Score*
1	95.00%	95.45%	95.00%	94.99%
2	100%	100%	100%	100%
3	100%	100%	100%	100%
4	100%	100%	100%	100%
5	100%	100%	100%	100%
6	95.00%	95.45%	95.00%	94.99%
7	100%	100%	100%	100%
8	95.00%	95.45%	95.00%	94.99%
9	95.00%	95.45%	95.00%	94.99%
10	100%	100%	100%	100%

Rata – Rata Akurasi	
Akurasi	98.00%
Precision	98.18%
Recall	98.00%
F1 – Score	97.99%

Evaluasi *SVM* dengan *Forward Selection* menunjukkan peningkatan akurasi, menandakan pemanfaatan fitur terpilih berhasil mengoptimalkan performa model. Validasi dilakukan melalui perhitungan manual *Confusion Matrix*.



Gambar 3.6 *Confusion Matrix*

Gambar 3.6 menunjukkan average confusion matrix dari model SVM dengan hasil yang sangat baik. Terdapat 98 true negative dan 98 true positive, serta hanya 2 kesalahan untuk masing-masing false positive dan false negative. Ini mengindikasikan bahwa model memiliki akurasi, presisi, dan recall yang sangat tinggi dalam membedakan kedua kelas.

$$ACCURACY \frac{98 + 98}{98 + 98 + 2 + 2} = 98\%$$

Rumus Akurasi menunjukkan perhitungan akurasi sebesar 98%, yang diperoleh dari jumlah prediksi benar (98 + 98) dibagi total semua prediksi (98 + 98 + 2 + 2). Nilai ini menunjukkan bahwa model memiliki tingkat ketepatan yang sangat tinggi dalam klasifikasi.

3.1.9 Perbandingan Hasil

Perbandingan performa model dilakukan terhadap tiga skenario: *SVM* tanpa penanganan ketidakseimbangan dan tanpa seleksi fitur, *SVM* setelah *ADASYN*, dan *SVM* dengan *ADASYN* serta *Forward Selection*. Model awal mencatat akurasi sebesar 70,75%, meningkat menjadi 96,38% setelah *ADASYN*, dan mencapai 98,00% setelah ditambahkan seleksi fitur. Secara keseluruhan, peningkatan akurasi mencapai 27,25 poin persentase. Hal ini menunjukkan pentingnya

integrasi teknik balancing dan seleksi fitur dalam meningkatkan kinerja klasifikasi.

Tabel 3.4 Perbandingan Hasil Akurasi Pengujian *SVM*

Fold	*SVM (Tanpa Forward Selection* dan *ADASYN*	*SVM setelah ADASYN* (Sebelum *Forward Selection*)	*SVM* (Sesudah *Forward Selection* dan *ADASYN*)	Perbandingan SVM + Setelah ADASYN	SVM setelah ADASYN + Setelah Forward Selection
1	69.40%	97.27%	95.27%	+27.87%	-2.06%
2	65.03%	93.44%	100%	+28.41%	+7.02%
3	71.43%	97.81%	100%	+26.38%	+2.23%
4	71.98%	97.80%	100%	+25.82%	+2.25%
5	72.53%	95.60%	100%	+23.07%	+4.60%
6	73.63%	96.70%	95.00%	+23.07%	-1.76%
7	73.63%	95.05%	100%	+21.42%	+5.21%
8	71.98%	96.15%	95.00%	+24.17%	-1.20%
9	70.33%	97.80%	95.00%	+27.47%	-2.86%
10	67.58%	96.15%	100%	+28.57%	+4.00%
Rata-rata Akurasi					
	70.75%	96.38%			98.00%

Penerapan *ADASYN* dan *Forward Selection* secara bertahap meningkatkan akurasi *SVM* dari 70,75% menjadi 98,00%. *ADASYN* menyumbang kenaikan 25,63%, sementara *Forward Selection* menambah 1,62%, menegaskan pentingnya penyeimbangan data dan seleksi fitur dalam meningkatkan performa klasifikasi.

3.1.9 Pembahasan

Penelitian ini menganalisis kinerja algoritma Support Vector Machine (SVM) dalam mengklasifikasikan tingkat keparahan kecelakaan lalu lintas di Kota Samarinda. Untuk meningkatkan akurasi, digunakan metode Forward Selection sebagai seleksi fitur dan ADASYN untuk menangani ketidakseimbangan data. Pengujian dilakukan secara bertahap: SVM standar, SVM dengan ADASYN, dan SVM yang dikombinasikan dengan Forward Selection. Pembahasan berikut menguraikan hasil analisis berdasarkan rumusan masalah yang telah ditetapkan.

a) Dua penelitian menunjukkan efektivitas teknik seleksi fitur dalam meningkatkan akurasi model klasifikasi. Penelitian pertama menggunakan algoritma *Support Vector Machine (SVM)* yang dikombinasikan dengan metode *Forward Selection* dan *ADASYN* untuk mengklasifikasikan tingkat keparahan kecelakaan lalu lintas di Kota Samarinda, menghasilkan peningkatan akurasi dari 70,75% menjadi 98% berkat pemilihan fitur penting seperti Cuaca, Jumlah Luka Ringan, dan Jumlah Luka Berat. Sementara itu, penelitian kedua oleh [21] menerapkan algoritma *Naïve Bayes* dengan *Backward Elimination* dalam klasifikasi prevalensi stunting pada anak balita, menghasilkan peningkatan akurasi dari 53,50% menjadi 92,54%. Kedua studi menegaskan bahwa seleksi fitur secara signifikan mampu meningkatkan kinerja klasifikasi dan efektivitas model dalam mengenali pola data yang kompleks maupun tidak seimbang.
b) Kombinasi metode *Forward Selection* dan *ADASYN* pada algoritma *SVM* terbukti meningkatkan akurasi klasifikasi secara signifikan, mencapai rata-rata 98%. Dibandingkan model awal yang hanya mencapai 70,75%, terjadi peningkatan lebih dari 25%. Hasil ini menunjukkan bahwa seleksi fitur dan penyeimbangan data efektif dalam meningkatkan performa model terhadap data kecelakaan yang kompleks dan tidak seimbang.

4. KESIMPULAN

Pada kesimpulan ini penerapan metode seleksi fitur *Forward Selection* pada data kecelakaan lalu lintas di Kota Samarinda menghasilkan empat fitur utama, yaitu Jumlah Meninggal Dunia, Jumlah Luka Berat, Jumlah Luka Ringan, dan Cuaca. Pengujian menunjukkan bahwa penerapan *ADASY*N dan Forward Selection secara bertahap meningkatkan akurasi model *SVM*, dari awalnya 70,75% menjadi 96,38% setelah *ADASYN*, dan meningkat lagi menjadi 98,00% dengan penambahan *Forward*

Selection. Secara keseluruhan, akurasi meningkat sebesar 25,63%, yang menunjukkan efektivitas kombinasi penyeimbangan data dan seleksi fitur dalam meningkatkan performa klasifikasi.

5. SARAN

Berdasarkan hasil penelitian yang telah dilakukan, disarankan agar penelitian selanjutnya mempertimbangkan penggunaan algoritma klasifikasi lain seperti *Random Forest* atau *XGBoost* untuk dibandingkan dengan kinerja *SVM*, serta mengeksplorasi metode seleksi fitur tambahan seperti *Recursive Feature Elimination (RFE)* guna meningkatkan akurasi model. Selain itu, penggunaan teknik *encoding* yang lebih tepat seperti *one-hot encoding* untuk data kategorikal *non-ordinal* juga perlu dipertimbangkan agar tidak terjadi misinterpretasi nilai. Validasi model dapat ditingkatkan dengan menggunakan variasi metode *cross-validation* yang lebih kompleks atau dataset dari wilayah lain untuk memastikan generalisasi hasil. Penelitian ini juga dapat dikembangkan lebih lanjut menjadi sistem peringatan dini kecelakaan lalu lintas berbasis *machine learning* yang dapat digunakan sebagai alat bantu pengambilan keputusan bagi pihak berwenang dalam meningkatkan keselamatan berkendara.

DAFTAR PUSTAKA

[1] BPS, "Jumlah Kecelakaan Lalu Lintas Menurut Bulan di Kota Samarinda, 2021-2023," Badan Pusat Statistik Kota Samarinda. https://samarindakota.bps.go.id/id/statistics-table/2/MjMyIzI=/jumlah-kecelakaan-lalu-lintas-menurut-bulan-di-kota-samarinda.html

[2] Adit Mustafa, "233 Kasus Laka Lantas Tercatat di Samarinda Sepanjang 2024 dan 28 Persen Meninggal Dunia," insitekaltim. https://insitekaltim.com/233-kasus-laka-lantas-tercatat-di-samarinda-sepanjang-2024-dan-28-persen-meninggal-dunia/

[3] A. Surya, "Fenomena Meningkatnya Angka Kecelakaan Lalu Lintas Di Kota Samarinda," Kumparan. https://kumparan.com/4_-alamuddin-surya/fenomena-meningkatnya-angka-kecelakaan-lalu-lintas-di-kota-samarinda-23Y4rhrPbiR/2

[4] 2023 Nabila Abda Salsabila et al., "Classification of The Severity 0f Traffic Accident Victims In The City of Samarinda Uses The K-Nearest Neighbor and Naive Bayes Algorithms," *Jurnal EKSPONENSIAL*, vol. 14, no. 2, pp. 99–106, 2023,http://jurnal.fmipa.unmul.ac.id/index.php/exponensial99

[5] M. Tiara *et al.*, "Pemanfaatan Algoritma ADASYN dan Support Vector Machine Dalam Meningkatkan Akurasi Prediksi Kanker Paru-Paru," vol. 8, no. 5, pp. 8773–8778, 2024.

[6] C. Chazar and B. Erawan, "Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine," *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, vol. 12, no. 1, pp. 67–80, 2020, doi: 10.37424/informasi.v12i1.48.

[7] E. Tiana and S. Wahyuni, "Hasil Analisis Teknik Data Mining dengan Metode Naive Bayes untuk Mendiagnosa Penyakit Kanker Payudara," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 1, no. 2, p. 130, 2020, doi: 10.30865/json.v1i2.1766.

[8] H. S. W. Hovi, A. Id Hadiana, and F. Rakhmat Umbara, "Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)," *Informatics and Digital Expert (INDEX)*, vol. 4, no. 1, pp. 40–45, 2022, doi: 10.36423/index.v4i1.895.

[9] B. Dimitrijevic, R. Asadi, and L. Spasovic, "Application of hybrid support vector Machine models in analysis of work zone crash injury severity," *Transp Res Interdiscip Perspect*, vol. 19, no. March, p. 100801, 2023, doi: 10.1016/j.trip.2023.100801.

[10] B. S. Susanti, R. H. Hirzi, S. A. Wulandhya, and S. H. Hastuti, "Analisis Klasifikasi Kecelakaan Lalu Lintas Lombok Timur Berdasarkan Tingkat Keparahan Korban Kecelakaan Menggunakan Metode Support Vector Machine dan Bootstrap Aggregating," vol. 1, no. 1, pp. 1–8, 2024.

[11] F. Rahmi, Ferra Yanuar, and Yudiantri Asdi, "Penggunaan Metode Support Vector Machine (SVM) dalam Mengidentifikasi Tingkat Keparahan Pada Kecelakaan Lalu Lintas," *Statistika*, vol. 24, no. 1, pp. 1–10, 2024, doi: 10.29313/statistika.v24i1.1690.

[12] Sopiatul Ulum, R. F. Alifa, P. Rizkika, and C. Rozikin, "Perbandingan Performa Algoritma KNN dan SVM dalam Klasifikasi Kelayakan Air Minum," *Generation Journal*, vol. 7, no. 2, pp. 141–146, 2023, doi: 10.29407/gj.v7i2.20270.

[13] M. R. Fanani, "Algoritma Naïve Bayes Berbasis Forward Selection Untuk Prediksi Bimbingan Konseling Siswa," *Jurnal DISPROTEK*, vol. 11, no. 1, pp. 13–22, 2020, doi: 10.34001/jdpt.v11i1.952.

[14] B. Wang, C. Zhang, Y. D. Wong, L. Hou, M. Zhang, and Y. Xiang, "Comparing Resampling Algorithms and Classifiers for Modeling Traffic Risk Prediction," *Int J Environ Res Public Health*, vol. 19, no. 20, 2022, doi: 10.3390/ijerph192013693.

[15] H. Marlisa, N. Satyahadewi, N. Imro'ah, and N. N. Debataraja, "Application of Adasyn Oversampling Technique on K-Nearest Neighbor

Algorithm," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 3, pp. 1829–1838, 2024, doi: 10.30598/barekengvol18iss3pp1829-1838.

[16] T. T. Wong and P. Y. Yeh, "Reliable Accuracy Estimates from k-Fold Cross Validation," *IEEE Trans Knowl Data Eng*, vol. 32, no. 8, pp. 1586–1594, 2020, doi: 10.1109/TKDE.2019.2912815.

[17] F. Latuconsina and M. S. N. Van Delsen, "Klasifikasi Menggunakan Metode Support Vector Machine (SVM) Multiclass pada Data Indeks Desa Membangun (IDM) di Provinsi Maluku," vol. 7, no. 2, pp. 380–395, 2024.

[18] Oryza Habibie Rahman, Gunawan Abdillah, and Agus Komarudin, "Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 17–23, 2021, doi: 10.29207/resti.v5i1.2700.

[19] Drajana, "Metode Support Vector Machine Dan Forward Selection Prediksi Pembayaran Pembelian Bahan Baku Kopra," *ILKOM Jurnal Ilmiah*, vol. 9, pp. 116–123, 2017.

[20] A. Mulyani, S. Khoerunisa, and D. Kurniadi, "Perbandingan Kinerja Algoritma KNN dan SVM Menggunakan SMOTE untuk Klasifikasi Penyakit Diabetes," vol. 14, pp. 25–34, 2025, doi: 10.22146/jnteti.v14i1.15198.

[21] M. Yunus, Muhammad Kunta Biddinika, and A. Fadlil, "Optimasi Algoritma Naïve Bayes Menggunakan Fitur Seleksi Backward Elimination untuk Klasifikasi Prevalensi Stunting," *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 3, no. 2, pp. 278–285, 2023, doi: 10.51454/decode.v3i2.188.