JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 13 No. 1, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i1.5890

IMPLEMENTASI METODE *NAIVE BAYES* UNTUK KLASIFIKASI KECELAKAAN LALU LINTAS DI KOTA SAMARINDA

Cindy Azra Salsabila[1], Fendy Yulianto[2], Taghfirul Azhima Yoga Siswa[3]

[1,2,3]Universitas Muhammadiyah Kalimantan Timur; Jl. Ir. H. Juanda No.15, Sidodadi, Kec. Samarinda Ulu, Kota Samarinda, Kalimantan Timur 75124; (0541) 748511

Received: 30 Desember 2024
Accepted: 14 Januari 2025
Published: 20 Januari 2025

Keywords:
Traffic Accidents;
Classification;
Naïve Bayes;
Accuracy.

Corespondent Email:
Fy415@umkt.ac.id

Abstrak. Kecelakaan lalu lintas merupakan permasalahan serius di Kota Samarinda yang dipengaruhi oleh berbagai faktor seperti kondisi cahaya, cuaca, kelas jalan, tipe jalan, kondisi permukaan jalan, kemiringan jalan, batas kecepatan di lokasi, dan status jalan berkontribusi terhadap tingkat kecelakaan lalu lintas. Dalam mengatasi permasalahan penentuan kecelakaan lalu lintas dapat menggunakan konsep klasifikasi dengan metode *Naive Bayes*. Data yang digunakan akan dibagi menjadi dua bagian dengan rasio 80:20 untuk pelatihan dan pengujian, serta divalidasi menggunakan *K-Fold Cross Validation* dengan K=12, kemudian didapatkan hasil akurasi sebesar 84%. Hasil ini menunjukkan bahwa metode *Naive Bayes* dapat digunakan untuk melakukan penentuan jenis kecelakaan lalu lintas yang ada di Kota Samarinda.

Abstract. *Traffic accidents are a serious problem in the city of Samarinda, influenced by various factors such as lighting conditions, weather, road class, road type, road surface conditions, road slope, speed limits at the location, and road status, which contribute to the traffic accident rate. In addressing the issue of determining traffic accidents, the classification concept using the Naive Bayes method can be employed. The data used will be divided into two parts with an 80:20 ratio for training and testing, and validated using K-Fold Cross Validation with K=12, resulting in an accuracy of 84%. This result indicates that the Naive Bayes method can be used to determine the types of traffic accidents in the city of Samarinda.*

1. PENDAHULUAN

Perkembangan pesat sebuah kota mendukung kemajuan infrastruktur transportasi serta layanan publik, namun peningkatan jumlah kendaraan maupun percepatan pembangunan yang tidak diimbangi dengan perbaikan kualitas jalan dan perhatian terhadap keselamatan dapat memicu terjadinya kecelakaan lalu lintas [1]. Kecelakaan lalu lintas merupakan peristiwa tak terduga di jalan yang dapat terjadi di mana dan kapan saja, menyebabkan cedera hingga korban jiwa hingga kerugian materi, kemudian dipengaruhi oleh berbagai faktor, seperti cuaca, perilaku pengemudi, serta kondisi jalan [2].

Kondisi jalan merupakan faktor yang sangat mempengaruhi kecelakaan, diperlukan analisis mendalam terhadap berbagai faktor yang berpengaruh guna mengidentifikasi penyebab utama kecelakaan lalu lintas [3]. Mengidentifikasi penyebab kecelakaan lalu lintas berdasarkan faktor-faktor yang

berkontribusi sangat penting, langkah penyelesaiannya dapat mencakup perbaikan infrastruktur jalan, pemasangan rambu lalu lintas, serta analisis penyebab kecelakaan menggunakan konsep sistem cerdas seperti sistem pendukung keputusan, sistem pakar, dan klasifikasi [4].

Klasifikasi merupakan proses analisis data untuk mengelompokkan atau memprediksi kategori berdasarkan pola serta karakteristik tertentu, dengan metode yang sering digunakan pada klasifikasi antara lain adalah *Decision Tree*, *Random Forest,* dan *Naïve Bayes* [5]. *Naïve Bayes* merupakan metode klasifikasi yang memanfaatkan probabilitas untuk membangun model prediksi, menghitung kemungkinan terjadinya suatu peristiwa dan menyesuaikan prediksi tersebut ketika terdapat data atau informasi tambahan [6].

Berdasarkan penelitian yang dilakukan oleh [7] menggunakan metode *Naive Bayes* untuk menganalisis data kecelakaan lalu lintas di Kota Metro Manila, didapatkan nilai akurasi model sebesar 70.03%. Penelitian selanjutnya dilakukan oleh [8] membandingkan metode *Neural Network*, *K-Nearest Neighbor*, *Decision Tree*, dan *Naive Bayes* untuk menganalisis data kecelakaan lalu lintas di Kota Calabarzon Filipina, didapatkan akurasi *Neural Network* sebesar 87.63%, *K-Nearest Neighbor* sebesar 82.01%, *Decision Tree* sebesar 74.20%, serta *Naive Bayes* sebesar 86.28%, menunjukkan bagaimana metode *Naive Bayes* menangani prediksi kecelakaan di lokasi tersebut.

Penelitian ini menggunakan metode *Naive Bayes* karena kemampuannya dalam menangani dataset kompleks dengan cepat, mampu mengolah data besar dengan fitur kategorikal, serta melatih model dan menghasilkan prediksi dengan cara yang sederhana [9]. Berdasarkan penelitian yang dilakukan oleh [10] dengan membandingkan klasifikasi kecelakaan lalu lintas menggunakan metode *Naive Bayes* dan *K-Nearest Neighbor* didapatkan hasil nilai akurasi *Naïve Bayes* mencapai 79.31% sedangkan *K-Nearest Neighbor* mencapai 75.86%.

Confusion Matrix merupakan informasi perbandingan antara hasil klasifikasi model dengan hasil sebenarnya, yang digunakan untuk menghitung *Accuracy*, *Precision*, *Recall*, serta *F1-Score* sebagai ukuran ketepatan prediksi, kemampuan mendeteksi kelas positif, kemudian keseimbangan antara *Precision* dan *Recall* [11]. *Confusion Matrix* digunakan untuk menilai kinerja model klasifikasi dengan memeriksa jumlah prediksi yang benar dan salah pada setiap kelas, memungkinkan analisis lebih mendalam terhadap kemampuan model dalam mengidentifikasi fitur masing-masing kelas [12].

2. TINJAUAN PUSTAKA

2.1 Objek Penelitian

Kecelakaan lalu lintas merupakan suatu peristiwa tidak dapat diprediksi atau disengaja oleh para pengguna jalan, melibatkan kendaraan, benda, yang sering kali menyebabkan beberapa konsekuensi [13]. Data yang digunakan pada penelitian ini adalah data sekunder kecelakaan lalu lintas di Kota Samarinda yang diperoleh dari Kepolisian Sektor Kota Samarinda, dengan total 1004 data kecelakaan dari tahun 2021 sampai 2024, dengan faktor-faktor kecelakaan lalu lintas yang sering terjadi di jalan beserta keterangannya dapat dilihat pada Tabel 1.

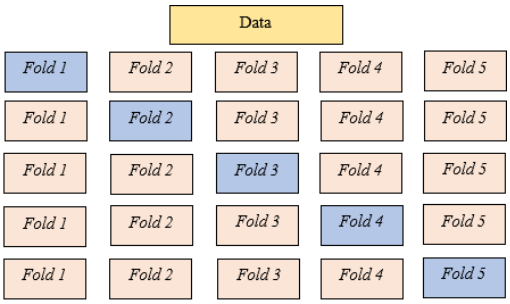
Tabel 1. Data penelitian

Atribut	Data
Kondisi cahaya	- Gelap - Terang
Cuaca	- Cerah - Hujan - Angin kencang - Hujan disertai angin kencang - Berawan
Kelas jalan	- I (Jalan besar) - II (Jalan sedang) - III (Jalan kecil)
Tipe jalan	- 1/1 (1 lajur/1 arah) - 2/2 tanpa batas median - 4/2 dengan batas median
Kondisi permukaan jalan	- Berombak - Baik - Licin - Berlubang - Basah
Kemiringan jalan	- Datar - Menanjak

Batas kecepatan di lokasi	- 20 - 30 - 40 - 50 - 60 - 80
Status jalan	- Jalan provinsi - Jalan kota - Jalan nasional - Jalan desa

2.2 K-Fold Cross Validation

K-Fold Cross Validation adalah metode yang menguji suatu algoritme dengan membagi data sampel secara acak, kemudian dikelompokkan sebanyak nilai K, di mana setiap bagian akan digunakan untuk proses pelatihan dan pengujian [14]. Adapun contoh proses *K-Fold Cross Validation* dapat dilihat pada Gambar 1.



Gambar 1. Contoh *K-Fold Cross Validation*

Berdasarkan Gambar 2.1 dalam contoh proses validasi *K-Fold Cross Validation* dilakukan dengan membagi dataset menjadi lima *Fold*, di mana setiap iterasi menggunakan empat *Fold* sebagai data pelatihan dan satu *Fold* sebagai data pengujian.

2.3 Klasifikasi

Klasifikasi merupakan pendekatan analisis yang digunakan untuk menentukan model atau fungsi dengan membedakan berbagai kelas dalam data dan memprediksi kelas objek berdasarkan atribut serta fitur yang dimiliki, berdasarkan suatu metode [15].

2.4 Binary Classification

Binary Classification merupakan proses membagi data ke dalam dua kelas berbeda, namun bila melebihi dua kelas, data akan diubah menjadi klasifikasi biner dengan memilih satu kelas sebagai kelas minoritas dan menggabungkan semua kelas lainnya menjadi kelas mayoritas [16].

2.5 Multi-Label Classification

Multi-Label Classification merupakan metode yang mempelajari suatu himpunan objek di mana setiap objek dapat termasuk lebih dari satu label, objek sering kali memiliki banyak karakteristik atau kategori yang relevan secara bersamaan [17].

2.6 Multi-Class Classification

Multi-Class Classification merupakan proses mengklasifikasikan data dengan menentukan kategori yang mencakup lebih dari dua kelas berbeda untuk sebuah objek, pada metode ini model dilatih membedakan antara berbagai kelas yang ada [18].

2.6.1 One-vs-One

One-vs-One merupakan strategi pendekatan yang menggunakan satu penggolong dalam setiap kelas, di mana setiap penggolong dilatih untuk membedakan antara dua kelas berbeda secara langsung [19].

2.6.2 One-vs-All

One-vs-All merupakan strategi pendekatan yang digunakan untuk menyelesaikan permasalahan *Multi-Class* dengan cara membandingkan setiap kelas secara terpisah dengan semua kelas lainnya [20].

2.7 Naïve Bayes

Naive Bayes merupakan metode analisis statisik untuk memprediksi probabilitas berbagai kelas dengan mengelompokkan permasalahan berdasarkan kesamaan dan perbedaan ciri-ciri atau fitur dari data [21]. Perhitungan untuk menghitung nilai Probabilitas *Prior* dapat menggunakan Persamaan 2.1.

$$P(Ci) = \frac{Si}{s} \quad (2.1)$$

Keterangan:
P(Ci) = Probabilitas *Prior*.
Si = Jumlah data dari kelas Ci.
s = Jumlah total data.

Menghitung nilai probabilitas *Likelihood* menggunakan *Mean* dapat menggunakan

Persamaan 2.2.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (2.2)$$

Keterangan:
μ = Rata-rata dari data.
n = Jumlah total data.
x_i = Nilai ke-i dari data.

Menghitung kuadrat *Mean* dapat menggunakan Persamaan 2.3.

$$(x_i - \mu)^2 \quad (2.3)$$

Keterangan:
x_i = Nilai individual dari data ke-i dalam kumpulan data.
μ = Nilai rata-rata dari data.
$(x_i - \mu)$= Selisih antara nilai individual dan rata-rata.

Menghitung nilai selisih kuadrat sebelum menghitung *Variance* dapat menggunakan Persamaan 2.4.

$$\sum_{i=1}^{N}(x_i - \mu)^2 \quad (2.4)$$

Keterangan:
N = Jumlah data.
x_i = Nilai data ke-i.
μ = Nilai rata-rata dari data.

Menghitung nilai Probabilitas *Likelihood* menggunakan *Variance* dapat menggunakan Persamaan 2.5.

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1} \quad (2.5)$$

Keterangan:
s^2 = *Variance* sampel.
x_i = Nilai individu dari dataset.
μ = *Mean* dari dataset.
n = Jumlah total data dari dataset.

Menghitung Probabilitas *Likelihood* dapat menggunakan Persamaan 2.6.

$$P(x_i|y) = \frac{N_{xiy}}{N_y} \quad (2.6)$$

Keterangan:
N_{xiy} = Jumlah kemunculan fitur ke-i dengan nilai x_i pada kelas y.
N_y = Jumlah total data pada kelas y.

Menghitung nilai *Evidence* dapat menggunakan Persamaan 2.7.

$$P(X) = \sum_{j=i}^{m} P(X|C_j) \times P(C_j) \quad (2.7)$$

Keterangan:
$P(X)$ = *Evidence.*
$\sum_{j=i}^{m}$ = Jumlah total kelas.
$P(X|C_j)$ = Nilai *Likelihood.*
$P(C_j)$ = Probabilitas *Prior* dari nilai C_j.

Menghitung Probabilitas *Posterior* dapat menggunakan Persamaan 2.8.

$$P(C_k|X) = \frac{P(X|C_k)\,.P(C_k)}{P(X)} \quad (2.8)$$

Keterangan:
$P(C_k|X)$ = Probabilitas *Posterior* dari kelas C_k terhadap data *X*.
$P(X|C_k)$ = Probabilitas data *X* terhadap kelas C_k atau *Likelihood*.
$P(C_k)$ = Probabilitas *Prior* dari kelas C_k sebelum mempertimbangkan data *X*.
$P(X)$ = Probabilitas total dari data *X* atau *Evidence.*

2.8 Evaluasi

Evaluasi dapat dilakukan untuk mengukur kinerja metode dalam menganalisis data kecelakaan lalu lintas dengan menggunakan *Confusion Matrix* untuk mencari nilai *Recall, Precision, Accuracy,* dan *F1-Score*.

2.8.1 Confusion Matrix

Confusion Matrix adalah alat evaluasi klasifikasi yang menggambarkan kinerja model dengan membandingkan prediksi model dan label sebenarnya [22].

2.8.2 Accuracy

Accuracy mengukur kedekatan nilai prediksi dengan nilai sebenarnya, berdasarkan perbandingan jumlah prediksi benar terhadap total prediksi [23]. Nilai *Accuracy* dapat menggunakan Persamaan 2.9.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.9)$$

Keterangan:
TP = *True Positive.*
TN = *True Negative.*
FP = *False Positive.*
FN = *False Negative.*

2.8.3 Precision

Precision mengukur akurasi sistem dalam memilih jawaban relevan, dihitung dari rasio jumlah jawaban relevan terhadap total jawaban.[24]. Nilai *Precision* dapat menggunakan Persamaan 2.10.

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2.10)$$

Keterangan:
TP = *True Positive.*
FP = *False Positive.*

2.8.4 Recall

Recall mengukur seberapa sering model memprediksi positif dengan benar saat kelas aktualnya positif, menunjukkan kemampuan model dalam mengidentifikasi kelas positif [25]. Nilai *Recall* dapat menggunakan Persamaan 2.11.

$$Recall = \frac{TP}{TP+FN} \quad (2.11)$$

Keterangan:
TP = *True Positive.*
FN = *False Negative.*

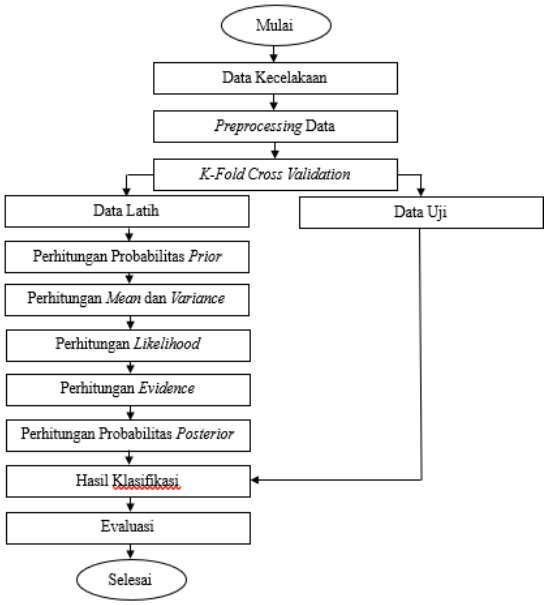
2.8.5 F1-Score

F1-Score adalah ukuran rata-rata harmonik antara Precision dan Recall, digunakan untuk mengevaluasi keseimbangan kinerja sistem. Nilai *F1-Score* dapat menggunakan Persamaan 2.12.

$$F1-Score = 2 \times \frac{Recall \times Presisi}{Recall + Presisi} \quad (2.12)$$

3. METODE PENELITIAN

Penelitian ini menjelaskan alur metode yang merupakan langkah-langkah dalam melakukan suatu penelitian. Alur metode dalam penelitian ini dapat dilihat pada Gambar 2.



Gambar 2. Alur Metode

Berdasarkan Gambar 2 alur metode dalam penelitian ini mencakup tahapan-tahapan yang dilakukan sebagai berikut:

1. Data Kecelakaan

Data kecelakaan lalu lintas yang digunakan dalam penelitian ini diperoleh dari Kepolisian Sektor Kota Samarinda periode 2021 hingga 2024.

2. *Preprocessing* Data

Preprocessing data pada penelitian ini merupakan proses pembersihan dan persiapan data sebelum digunakan untuk pelatihan model.

3. K-Fold Cross Validation

Data diuji dengan membagi dataset menjadi k lipatan, satu lipatan sebagai data uji, dan sisanya untuk pelatihan, diulang hingga k kali.

4. Pembagian Data

Data dibagi menjadi dua subset yaitu data latih dan data uji, di mana data latih digunakan untuk melatih model sementara data uji digunakan menguji model setelah dilatih.

5. Probabilitas *Prior*

Probabilitas *Prior* dihitung dari frekuensi tiap kelas dalam dataset.

6. Perhitungan *Likelihood*

Perhitungan *Mean* dan *Variance* dilakukan sebelum tahap *Likelihood*, yang mengukur probabilitas kemunculan fitur dalam sebuah kelas berdasarkan rasio jumlah kemunculan kelas terhadap total data dalam kelas.

7. Perhitungan *Evidence*

Evidence merupakan data yang digunakan untuk menghitung probabilitas dan menentukan

kemungkinan suatu data termasuk dalam kelas tertentu.

8. Probabilitas *Posterior*

Probabilitas *Posterior* merupakan probabilitas mengukur seberapa besar kemungkinan suatu kelas terjadi setelah mempertimbangkan *Evidence* yang telah diamati.

9. *Hasil Klasifikasi*

Proses ini membandingkan nilai Probabilitas *Posterior* tiap kelas, dengan kelas bernilai tertinggi dianggap paling mungkin, memastikan data diklasifikasikan tepat dan meningkatkan akurasi model.

10. *Evaluasi*

Evaluasi merupakan analisis mendalam terhadap hasil prediksi yang dihasilkan oleh model untuk memastikan bahwa model tersebut berfungsi dengan baik

4. HASIL DAN PEMBAHASAN

4.1. Preprocessing Data

Preprocessing data dalam penelitian ini meliputi pembersihan data dengan penanganan nilai yang hilang, identifikasi kesalahan, pemilihan fitur, dan transformasi data. Transformasi data dalam penelitian ini menggunakan label *Encoding* untuk mengubah data kategorikal menjadi format numerik, sebagaimana ditunjukkan pada Gambar 3.

	Kondisi_Cahaya	Cuaca	Kelas_Jalan	Tipe_Jalan	Kondisi_Permukaan_Jalan	Batas_Kecepatan	Kemiringan_Jalan	Status_Jalan	Label
0	0	0	1	0	1	0	0	1	0
1	0	0	1	0	1	0	1	1	0
2	1	0	1	1	0	0	0	1	0
3	0	0	0	0	0	1	0	0	0
4	0	0	0	0	1	0	1	0	0
5	0	0	1	1	1	0	0	1	0
6	1	1	1	1	1	1	1	1	2
7	0	0	1	0	0	0	0	1	0
8	0	0	1	1	0	0	0	0	0
9	0	0	1	1	0	1	0	1	0

Gambar 3. *Preprocessing* Data

Berdasarkan Gambar 3 label “0” menyatakan bahwa faktor tidak berpengaruh sedangkan label “1” menunjukkan adanya pengaruh dengan tingkat kecelakaan dikelompokkan sebagai “0” kelas ringan, “1” sedang, dan “2” berat.

4.2. Implementasi K-Fold Cross Validation

Data dibagi menjadi beberapa subset dan pada implementasi ini menggunakan nilai K=7 data dibagi menjadi tujuh *Fold* untuk pelatihan serta pengujian model, implementasi *K-Fold Cross Validation* dapat dilihat pada Gambar 4.

	Fold	Index	Actual (Biner)	Predicted (Biner)
0	1	0	1	1
1	1	1	0	1
2	1	2	1	1
3	1	3	1	0
4	1	4	0	0
...	...	...	...	...
7023	7	999	1	1
7024	7	1000	1	1
7025	7	1001	1	1
7026	7	1002	1	1
7027	7	1003	1	1

Gambar 4. *K-Fold Cross Validation*

4.3. Pembagian Data

Data dibagi menjadi dua bagian, dengan 80:20 atau 80% data latih dan 20%, pembagian data dapat dilihat pada Gambar 5.

```
Jumlah data latih: 803
Jumlah data uji: 201
```

Gambar 5. Pembagian Data

4.4. Probabilitas Prior

Probabilitas *Prior* merupakan estimasi awal kemungkinan setiap kelas sebelum mempertimbangkan fitur-fitur, Perhitungan Probabilitas *Prior* dapat dilihat pada Gambar 6.

```
Probabilitas Prior:
Label
0    0.585657
1    0.342629
2    0.071713
```

Gambar 6. Probabilitas *Prior*

4.5. Mean dan Variance

Perhitungan *Mean* dan *Variance* dilakukan pada setiap fitur dalam dataset berdasarkan masing-masing kelas, *Mean* digunakan untuk nilai tengah distribusi data dalam setiap kelas, sedangkan *Variance* mengukur sejauh mana data menyebar melalui nilai-nilai tersebut.

4.6. Probabilitas Likelihood

Probabilitas *Likelihood* adalah ukuran yang menentukan seberapa sering suatu fitur muncul dalam sebuah kelas tertentu dengan membandingkan frekuensi kemunculannya terhadap total data dalam kelas tersebut.

4.7. Evidence

Evidence merupakan seberapa besar kemungkinan suatu data termasuk dalam kelas

tertentu dengan mempertimbangkan fitur-fiturnya, hasil perhitungan *Evidence* dapat dilihat pada Gambar 7.

```
Evidence P(X): 9.907337049719199e-21
```

Gambar 7. *Evidence*

4.8. Probabilitas Posterior

Probabilitas *Posterior* merupakan kemungkinan akhir sebuah data termasuk dalam kelas tertentu setelah mempertimbangkan semua fitur, hasil Probabilitas *Posterior* dapat dilihat pada Gambar 8.

```
Probabilitas Posterior: {'Ringan': 0.5494, 'Sedang': 0.4216, 'Berat': 0.029}
Prediksi Kelas: Ringan

Probabilitas Posterior: {'Ringan': 0.4161, 'Sedang': 0.5101, 'Berat': 0.0738}
Prediksi Kelas: Sedang

Probabilitas Posterior: {'Ringan': 0.6058, 'Sedang': 0.3378, 'Berat': 0.0564}
Prediksi Kelas: Ringan

Probabilitas Posterior: {'Ringan': 0.2963, 'Sedang': 0.7037, 'Berat': 0.0}
Prediksi Kelas: Sedang

Probabilitas Posterior: {'Ringan': 0.504, 'Sedang': 0.265, 'Berat': 0.231}
Prediksi Kelas: Ringan
```

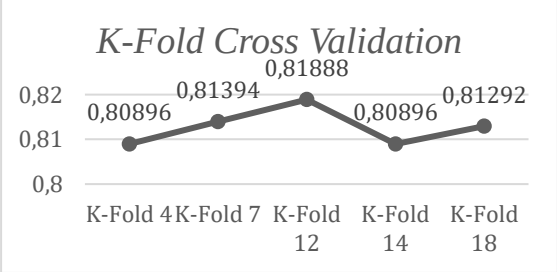
Gambar 8. Probabilitas *Posterior*

4.9. Data Kelas Aktual, dan Prediksi

Data kelas aktual merupakan label sebenarnya dari setiap data uji, sedangkan data prediksi adalah hasil klasifikasi yang dihasilkan oleh model.

4.10. K-Fold Cross Validation

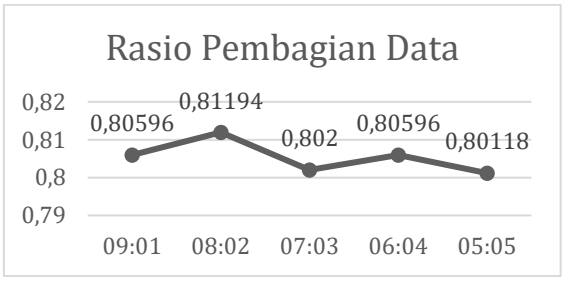
K-Fold Cross Validation dalam penelitian ini melibatkan lima iterasi dengan nilai K yang berbeda K=4,7,12,14,18, di mana setiap nilai K mewakili jumlah lipatan dalam proses validasi dan evaluasi dapat dilihat pada Gambar 9.



Gambar 9. *K-Fold Cross Validation*

4.11. Rasio Pembagian Data

Rasio pembagian data dilakukan dengan berbagai rasio pembagian data untuk mengevaluasi kinerjanya, rasio yang digunakan meliputi 90:10, 80:20, 70:30, 60:40, dan 50:50, hasil rasio pembagian data dapat dilihat pada Gambar 10.



Gambar 10. Rasio Pembagian Data

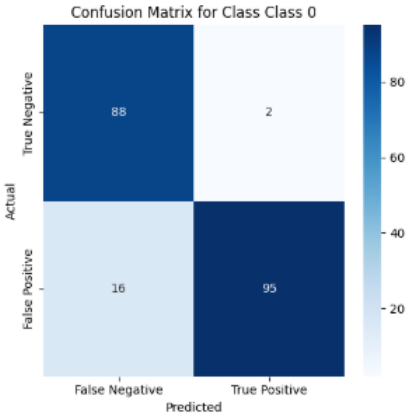
4.12. Pengujian Awal

Pengujian awal metode *Naïve Bayes* dengan *K-Fold Cross Validation* dan rasio pembagian data memberikan gambaran awal tentang kinerja model terhadap data yang ada serta mendukung evaluasi lebih lanjut, seperti yang dapat dilihat pada Gambar 11.

	precision	recall	f1-score	support
0	0.99	0.88	0.93	126
1	0.73	0.62	0.67	58
2	0.42	1.00	0.60	17
accuracy			0.82	201
macro avg	0.72	0.83	0.73	201
weighted avg	0.87	0.82	0.83	201

Gambar 11. Pengujian Awal

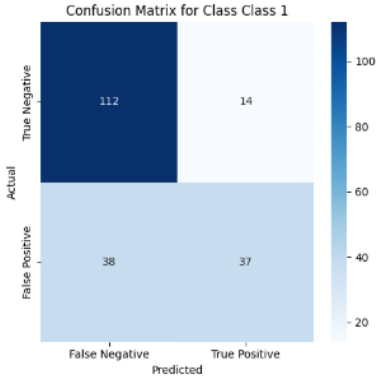
Adapun *Confusion Matrix* pengujian awal untuk kelas 0 dalam menentukan tingkat kecelakaan lalu lintas dapat dilihat pada Gambar 12.



Gambar 12. *Confusion Matrix* untuk kelas 0

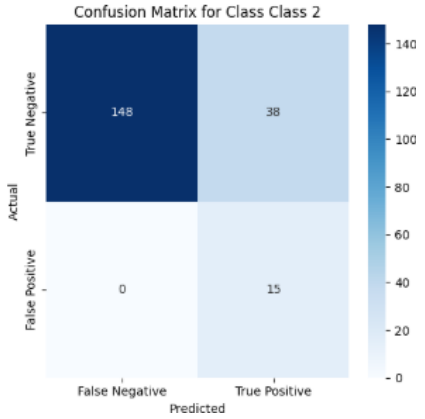
Adapun *Confusion Matrix* pengujian awal untuk kelas 1 dalam menentukan tingkat

kecelakaan lalu lintas dapat dilihat pada Gambar 13.



Gambar 13. *Confusion Matrix* untuk kelas 1

Adapun *Confusion Matrix* pengujian awal untuk kelas 2 dalam menentukan tingkat kecelakaan lalu lintas dapat dilihat pada Gambar 14.



Gambar 14. *Confusion Matrix* untuk kelas 2

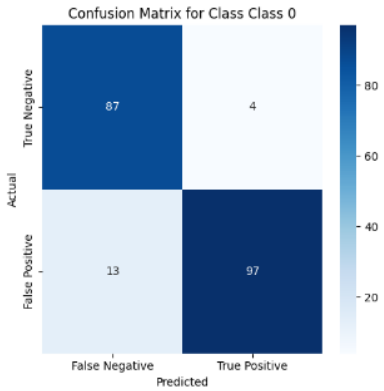
4.13. Pengujian Akhir

Pengujian akhir dilakukan untuk mengevaluasi kinerja model dengan rasio data 80:20 dan *K-Fold Cross Validation* optimal (K=12), pengujian akhir dapat dilihat pada Gambar 15.

	precision	recall	f1-score	support
0	0.98	0.93	0.95	120
1	0.86	0.65	0.74	66
2	0.39	1.00	0.57	15
accuracy			0.84	201
macro avg	0.75	0.86	0.75	201
weighted avg	0.90	0.84	0.85	201

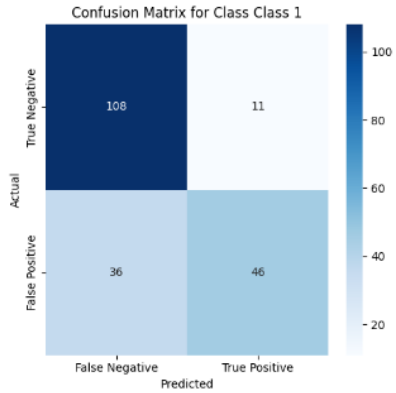
Gambar 15. Pengujian Akhir

Adapun *Confusion Matrix* pengujian awal untuk kelas 0 dalam menentukan tingkat kecelakaan lalu lintas dapat dilihat pada Gambar 16.



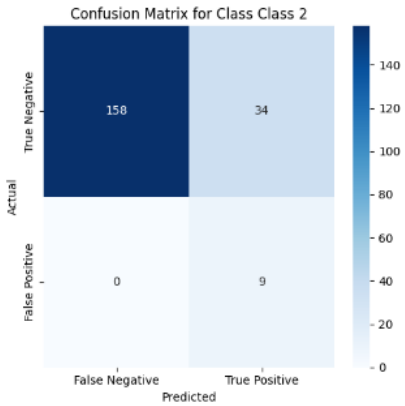
Gambar 16. *Confusion Matrix* untuk kelas 0

Adapun *Confusion Matrix* pengujian awal untuk kelas 1 dalam menentukan tingkat kecelakaan lalu lintas dapat dilihat pada Gambar 17.



Gambar 17. *Confusion Matrix* untuk kelas 1

Adapun *Confusion Matrix* pengujian awal untuk kelas 2 dalam menentukan tingkat kecelakaan lalu lintas dapat dilihat pada Gambar 18.



Gambar 18. *Confusion Matrix* untuk kelas 2

4.14. Pembahasan

Penelitian ini menggunakan data kecelakaan lalu lintas di Kota Samarinda sebanyak 1004 yang diperoleh dari Kepolisian Sektor Kota Samarinda melalui pengumpulan data sekunder berdasarkan hasil observasi dan wawancara dengan menerapkan model *Naïve Bayes* untuk klasifikasi tingkat kecelakaan, melibatkan perhitungan Probabilitas *Prior*, *Likelihood*, *Evidence* dan Probabilitas *Posterior*. Pengujian ini menggunakan teknik *K-Fold Cross Validation* dengan K=12 dan rasio pembagian data 20% yang dilakukan lima kali untuk memastikan model diuji secara menyeluruh dan menghasilkan akurasi tertinggi sebesar 84% serta memungkinkan penilaian mendalam terhadap kinerja model pada berbagai subset data.

5. KESIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan maka didapatkan kesimpulan sebagai berikut:

a. Klasifikasi tingkat kecelakaan lalu lintas di Kota Samarinda dilakukan menggunakan metode *Naïve Bayes* melibatkan tahapan perhitungan seperti Probabilitas *Prior*, Probabilitas *Likelihood*, *Evidence* dan Probabilitas *Posterior*. Probabilitas *Posterior* ini berperan penting dalam menentukan dengan tepat ke kelas mana setiap kejadian kecelakaan diklasifikasikan sehingga memastikan bahwa model dapat memberikan ketepatan dalam melakukan perhitungan probabilitas.
b. Terdapat peningkatan dalam nilai akurasi pengujian dengan *Accuracy* pengujian awal sebesar 82% dan *Accuracy* pengujian akhir meningkat menjadi 84% setelah menerapkan nilai *K-Fold* dan rasio pembagian data tertinggi. Peningkatan ini menunjukkan bahwa nilai *K-Fold* dan rasio pembagian data yang diterapkan berhasil meningkatkan akurasi dalam klasifikasi tingkat kecelakaan, hal ini menunjukkan pentingnya pemilihan nilai yang tepat dalam meningkatkan akurasi model dalam klasifikasi data.

DAFTAR PUSTAKA

1] S. Sarjana, "Urban Public Transportation Perspective in Meta-Analysis Study," *Media Komun. Tek. Sipil*, vol. 27, no. 2, pp. 277–287, 2021, doi: 10.14710/mkts.v27i2.40635.

[2] F. R. M and E. Widowati, "Kecelakaan Lalu Lintas Jalan Tol Ruas Batang-Semarang Berdasarkan Karakteristik Faktor Penyebab Kecelakaan Tahun 2019," *Indones. J. Public Heal. Nutr.*, vol. 1, no. 2, pp. 214–222, 2021, [Online]. Available: http://journal.unnes.ac.id/sju/index.php/IJPHN

[3] A. Franseda, W. Kurniawan, S. Anggraeni, and W. Gata, "Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 3, p. 282, 2020, doi: 10.26418/justin.v8i3.40982.

[4] Ardi Ramdani, Christian Dwi Sofyan, Fauzi Ramdani, Muhamad Fauzi Arya Tama, and Muhammad Angga Rachmatsyah, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Masyarakat Dalam Menerima Bantuan Sosial," *J. Ilm. Sist. Inf.*, vol. 1, no. 2, pp. 39–47, 2022, doi: 10.51903/juisi.v1i2.363.

[5] F. Yulianto, R. Arifando, and A. A. Supianto, "Improved Eliminate Particle Swarm Optimization on Support Vector Machine for Freshwater Fish Classification," *Proc. 2019 4th Int. Conf. Sustain. Inf. Eng. Technol. SIET 2019*, pp. 38–43, 2019, doi: 10.1109/SIET48054.2019.8986119.

[6] A. Karima and T. A. Y. Siswa, "Prediksi Kinerja Mahasiswa Dalam Perkuliahan Berbasis Learning Management System Menggunakan Algoritma Naïve Bayes," *Progresif J. Ilm. Komput.*, vol. 18, no. 2, p. 211, 2022, doi: 10.35889/progresif.v18i2.922.

[7] M. Libnao, M. Misula, C. Andres, J. Mariñas, and A. Fabregas, "Traffic incident prediction and classification system using naïve bayes algorithm," *Procedia Comput. Sci.*, vol. 227, pp. 316–325, 2023, doi: 10.1016/j.procs.2023.10.530.

[8] K. A. R. Torres and J. R. Asor, "Machine learning approach on road accidents analysis in Calabarzon, Philippines: An input to road safety management," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 24, no. 2, pp. 993–1000, 2021, doi: 10.11591/ijeecs.v24.i2.pp993-1000.

[9] A. Wisnu Saputra, A. Irma Purnamasari, and I. Ali, "Implementasi Algoritma Naïve Bayes Untuk Memprediksi Kualitas Air Yang Dapat Di Konsumsi," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 133–140, 2024, doi: 10.36040/jati.v8i1.8292.

[10] K. Tingkat *et al.*, "Classification of The Severity 0f Traffic Accident Victims In The City of Samarinda Uses The K-Nearest Neighbor and Naive Bayes Algorithms," *J.*

EKSPONENSIAL, vol. 14, no. 2, pp. 99–106, 2023, [Online]. Available: http://jurnal.fmipa.unmul.ac.id/index.php/eksponensial99

[11] B. Karnadi and T. Handhayani, "Klasifikasi Jenis Buah dengan Menggunakan Metode," pp. 35–42, 2024, doi: 10.30864/eksplora.v14i1.1067.

[12] R. Rousyati, A. N. Rais, N. Hasan, R. F. Amir, and W. Warjiyono, "Komparasi Adaboost dan Bagging Dengan Naïve Bayes Pada Dataset Bank Direct Marketing," *Bianglala Inform.*, vol. 9, no. 1, pp. 12–16, 2021, doi: 10.31294/bi.v9i1.9890.

[13] N. M. L. Dewi, "Implementasi Alternative Dispute Resolution Dalam Penyelesaian Kasus Kecelakaan Lalu Lintas Warga Negara Asing," *Semin. Nas. INOBALI Inov. Baru dalam Penelit. Sains, Teknol. dan Hum.*, no. 15, pp. 679–686, 2019.

[14] R. R. R. Arisandi, B. Warsito, and A. R. Hakim, "Aplikasi Naïve Bayes Classifier (Nbc) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation," *J. Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, doi: 10.14710/j.gauss.v11i1.33991.

[15] A. P. Wibawa *et al.*, "Naïve Bayes Classifier for Journal Quartile Classification," *Int. J. Recent Contrib. from Eng. Sci. IT*, vol. 7, no. 2, p. 91, 2019, doi: 10.3991/ijes.v7i2.10659.

[16] T. Kim and J. S. Lee, "Exponential Loss Minimization for Learning Weighted Naive Bayes Classifiers," *IEEE Access*, vol. 10, pp. 22724–22736, 2022, doi: 10.1109/ACCESS.2022.3155231.

[17] G. Mustafa, M. Usman, L. Yu, M. T. afzal, M. Sulaiman, and A. Shahid, "Multi-label classification of research articles using Word2Vec and identification of similarity threshold," *Sci. Rep.*, vol. 11, no. 1, pp. 1–20, 2021, doi: 10.1038/s41598-021-01460-7.

[18] K. Shah and D. K. S. Patel, "Study of Multiclass Classification Techniques," *Iarjset*, vol. 11, no. 2, pp. 131–137, 2024, doi: 10.17148/iarjset.2024.11220.

[19] Z. Alhaq, A. Mustopa, S. Mulyatun, and J. D. Santoso, "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter," *J. Inf. Syst. Manag.*, vol. 3, no. 2, pp. 44–49, 2021, doi: 10.24076/joism.2021v3i2.558.

[20] R. S. Tantika and A. Kudus, "Penggunaan Metode Support Vector Machine Klasifikasi Multiclass pada Data Pasien Penyakit Tiroid," *Bandung Conf. Ser. Stat.*, vol. 2, no. 2, pp. 159–166, 2022, doi: 10.29313/bcss.v2i2.3590.

[21] S. Lestari, A. Akmaludin, and M. Badrul, "Implementasi Klasifikasi Naive Bayes Untuk Prediksi Kelayakan Pemberian Pinjaman Pada Koperasi Anugerah Bintang Cemerlang," *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 7, no. 1, pp. 8–16, 2020, doi: 10.30656/prosisko.v7i1.2129.

[22] M. R. RahmaPutri, D. S. Y. Kartika, and S. F. A. Wati, "Klasifikasi Tingkat Kepuasan Pelanggan Sat & Sun : the Almeaty Service Menggunakan Naive Bayesklasifikasi Tingkat Kepuasan Pelanggan Sat & Sun : the Almeaty Service Menggunakan Naive Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4844.

[23] A. A. Elrahman, A. A. Elrahman, M. R. Riad, and M. M. Abdelgwad, "Predicting Adults ' Income using Naive Bayes Classi er Predicting Adults ' Income using Naive Bayes Classifier," pp. 0–7, 2024.

[24] G. M. C. Batubara, A. Desiani, and A. Amran, "Klasifikasi Jamur Beracun Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbors," *J. Ilmu Komput. dan Inform.*, vol. 3, no. 1, pp. 33–42, 2023, doi: 10.54082/jiki.68.

[25] J. V. Wie and M. Siddik, "Penerapan Metode Naïve Bayes Dalam Mengklasifikasi Tingkat Obesitas Pada Pria," *JOISIE J. Inf. Syst. Informatics Eng.*, vol. 6, no. Desember, pp. 69–77, 2022, [Online]. Available: https://www.kaggle.com/,