

# IMPLEMENTASI TEXT MINING DENGAN NAÏVE BAYES UNTUK ANALISIS SENTIMEN LAYANAN KEPOLISIAN SELAMA MUDIK2024

Fadhilah Fauzan Rachman<sup>1\*</sup>, Arief Wibowo<sup>2</sup>

<sup>1,2</sup> Teknik Informatika, Fakultas Teknik Informatika, Universitas Budi Luhur, Jakarta, Indonesia

Email: <sup>1\*</sup>2011501703@student.budiluhur.ac.id, <sup>2</sup>arief.wibowo@budiluhur.ac.id

(\* : corresponding author)

**Abstrak-**Mudik adalah tradisi tahunan bagi warga Indonesia yang kembali ke kampung halaman menjelang Idul Fitri untuk berkumpul bersama keluarga. Pada mudik tahun 2024, pemerintah dan Kepolisian Republik Indonesia telah meningkatkan keamanan dengan menambah personel di titik rawan kemacetan dan kecelakaan. Analisis opini masyarakat di platform X menunjukkan dua arah tanggapan terhadap layanan kepolisian selama mudik 2024. Banyak pengguna mengapresiasi responsivitas dan peningkatan keamanan, namun terdapat juga keluhan tentang beberapa area yang perlu perbaikan. Penelitian ini menggunakan metode Text Mining dan Algoritma Naïve Bayes untuk menganalisis sentimen pengguna X Indonesia terkait layanan kepolisian selama mudik 2024. Dataset diperoleh dari X harvest dengan rentang waktu 7-14 April 2024, terdiri dari 1.176 data dengan kata kunci "info mudik" dan "kabar mudik". Hasil analisis menunjukkan akurasi 63%, presisi 70%, recall 69%, dan f1-score 62%. Sentimen positif mencapai 50%, negatif 16%, dan netral 34%, menunjukkan mayoritas pengguna memiliki pandangan positif terhadap layanan kepolisian selama mudik 2024. Penelitian ini memberikan wawasan tentang persepsi publik terhadap upaya keamanan mudik dan dapat menjadi dasar bagi perbaikan layanan di masa mendatang. Maka dari itu Kontribusi penelitian ini terletak pada pembuatan sistem aplikasi berbasis web diharapkan dapat membantu pihak kepolisian untuk merespons dan menindaklanjuti setiap peristiwa yang terjadi selama pelaksanaan mudik selanjutnya, memastikan bahwa analisis sentimen yang akurat dapat menjadi alat bantu penting dalam pengambilan keputusan dan peningkatan pelayanan publik.

**Kata Kunci:** Analisis Sentimen, Polisi, Naïve Bayes

## IMPLEMENTATION OF TEXT MINING WITH NAIVE BAYES FOR POLICE SENTIMENT ANALYSIS OF SERVICES DURING HOMECOMING

**Abstract-**Mudik is an annual tradition for Indonesians returning to their hometowns before Eid al-Fitr to celebrate with family. In 2024, the government and the Indonesian National Police enhanced security by deploying additional personnel at traffic congestion and accident-prone areas. Public opinion analysis on the X platform revealed mixed reactions to police services during the 2024 mudik period. Many users appreciated the responsiveness and increased security, while some highlighted areas needing improvement. This study employs Text Mining and the Naïve Bayes algorithm to analyze the sentiment of Indonesian X users regarding police services during the 2024 mudik period. The dataset was obtained using X harvest from April 7 to April 14, 2024, comprising 1,176 records with keywords "info mudik" and "kabar mudik." The analysis results showed an accuracy of 63%, precision of 70%, recall of 69%, and an f1-score of 62%. Positive sentiment accounted for 50%, negative sentiment 16%, and neutral sentiment 34%, indicating that the majority of users held a positive view of police services during the 2024 mudik period. This study provides insights into public perceptions of mudik security efforts and can serve as a basis for improving future services.

**Keywords:** Sentiment Analysis, Police, Naïve Bayes

### 1. PENDAHULUAN

Mudik merupakan kebiasaan tahunan yang dilakukan oleh banyak warga Indonesia yang tinggal di kota besar untuk kembali ke tempat asal mereka sebelum hari raya keagamaan, khususnya Idul Fitri. Tradisi ini bertujuan untuk berkumpul dan merayakan hari besar tersebut bersama keluarga. Tradisi ini tidak hanya penting secara budaya tetapi juga memiliki pengaruh ekonomi, mempengaruhi berbagai aspek kehidupan masyarakat di perkotaandan pedesaan.

Tradisi mudik telah ada sejak lama dan berkembang seiring dengan perkembangan demografi dan ekonomi di Indonesia. Mudik menjadi lebih populer dan terstruktur seiring dengan peningkatan jumlah penduduk yang merantau mencari penghidupan di kota-kota besar seperti Jakarta, Bandung, atau Surabaya. Di era modern, pemerintah dan berbagai sektor terkait mulai menyelenggarakan fasilitas dan layanan khusus untuk mendukung kegiatan mudik, mengingat volume perjalanan yang meningkat signifikan dan tantangan logistik yang

ditimbulkannya. Sejarah mudik juga tidak lepas dari peningkatan infrastruktur transportasi seperti jalan raya, kereta api, dan layanan bus antar kota.

Pada mudik tahun 2024, pemerintah bersama dengan Kepolisian Republik Indonesia telah menyiapkan serangkaian langkah untuk meningkatkan keamanan dan kenyamanan bagi pemudik. Terdapat penambahan personil keamanan di titik-titik *Rawan* kemacetan dan kecelakaan, serta peningkatan fasilitas posko mudik yang dilengkapi dengan pelayanan medis dan perbaikan kendaraan darurat. Kepolisian juga mengimplementasikan sistem *oneway* di beberapa ruas jalan utama pada puncak arus mudik dan arus balik untuk memperlancar lalu lintas. Selain itu, aplikasi dan media sosial digunakan untuk memberikan informasi real-time mengenai kondisi jalan, cuaca, dan tips perjalanan selama mudik, memastikan pengalaman mudik yang lebih aman dan menyenangkan bagi masyarakat.

Berdasarkan analisis terhadap opini masyarakat pengguna X di Indonesia, terlihat ada dua arah tanggapan terhadap layanan kepolisian selama mudik 2024. Banyak pengguna mengapresiasi kepolisian atas responsivitas dan keamanan yang ditingkatkan selama periode mudik, terutama dalam pengaturan lalu lintas dan penanganan insiden kecil dengan cepat dan efisien. Selain itu, berkat peningkatan infrastruktur yang telah dilakukan pemerintah, seperti pembangunan jalan tol baru dan perbaikan sistem transportasi umum, dapat mengurangi kemacetan yang sering terjadi selama musim mudik. Namun, di sisi lain, terdapat pula keluhan tentang perlunya perbaikan dalam beberapa area. Untuk mengatasi ini, kepolisian dan lembaga terkait telah menyiapkan skema khusus untuk menghadapi lonjakan jumlah pemudik dan menerapkan strategi yang lebih baik dalam manajemen sumber daya dan komunikasi kepolisian untuk mengoptimalkan kepuasan publik. Oleh karena itu, Analisis Sentimen bisa dilakukan untuk menentukan apakah banyak dari opini tersebut bersifat positif, negatif, atau netral. Untuk melakukan Analisis Sentimen dengan menggunakan *Text Mining* untuk pengumpulan opini-opini tersebut, terdapat berbagai Algoritma yang dapat digunakan. Beberapa Algoritma yang sering digunakan dalam text mining untuk tujuan ini meliputi KNN, *Naive Bayes*, SVM, dan beberapa Algoritma lainnya.

Penelitian sebelumnya yang fokus pada Mobil Listrik memberikan dasar pemikiran yang berbeda dibandingkan dengan studi saat ini yang mengeksplorasi Layanan Kepolisian Dalam Masa Mudik 2024, menggunakan tweet-harvest untuk crawling dataset, sebuah metode yang belum dijelaskan secara rinci dan belum dirujuk dalam jurnal terkait dalam pendahuluan. Selain itu, penelitian ini mengimplementasikan pendekatan labeling otomatis menggunakan kamus lexicon, berbeda dari pendekatan manual yang digunakan pada penelitian sebelumnya, memberikan keunggulan dalam efisiensi dan ketepatan analisis sentimen yang belum disoroti secara memadai sebelumnya. Kelebihan ini menggarisbawahi pentingnya metode yang dipilih untuk mengatasi tantangan penelitian yang unik dan spesifik terkait dengan topik yang berbeda dan kompleksitas data yang berkembang.

Pembahasan penelitian ini akan berfokus pada Analisis Sentimen Masyarakat Terhadap Layanan Kepolisian Terkait Mudik 2024 Pada Media Sosial X Menggunakan Algoritma *Naive Bayes*. Penelitian ini diharapkan dapat memberikan informasi opini publik terkait Layanan Kepolisian Pada Masa Mudik 2024.

Kontribusi penelitian ini terletak pada pembuatan sistem aplikasi berbasis web diharapkan dapat membantu pihak kepolisian untuk merespons dan menindaklanjuti setiap peristiwa yang terjadi selama pelaksanaan mudik selanjutnya, memastikan bahwa analisis sentimen yang akurat dapat menjadi alat bantu penting dalam pengambilan keputusan dan peningkatan pelayanan publik.

## 2. METODE PENELITIAN

### 2.1. Pengumpulan Data

Di langkah ini, peneliti mengumpulkan informasi dari X menggunakan kata kunci "info mudik" dan "kabar mudik" selama periode tiga hari sebelum dan setelah Hari Raya Idul Fitri 2024 untuk mempelajari komunikasi seputar mudik dalam media sosial. Pemilihan kata kunci ini didasarkan pada penggunaannya yang luas dalam dialog terkait mudik, di mana individu seringkali berbagi dan mencari informasi mengenai situasi lalu lintas dan saran perjalanan dengan menggunakan kata kata "info mudik", "kabar mudik, ataupun *hashtag* "#kabarmudik dan "#infomudik". Hal ini memfasilitasi peneliti dalam memahami dan menganalisis tren serta pola yang muncul dalam diskusi mudik, yang sangat berguna untuk perencanaan dan pengelolaan mudik yang efektif. Dengan mengumpulkan 1.176 tweet dan mengonversinya ke format xlsx, riset ini bertujuan untuk memberikan analisis komprehensif tentang bagaimana mudik diinterpretasikan dan dibahas dalam media sosial.

### 2.2. Preprocessing

Pada tahap *Preprocessing*, dilakukan berbagai proses untuk menyaring, membuang, dan memperbaiki kata-kata. Tujuannya adalah menghasilkan teks yang lebih terstruktur dari data tweet, yang dikenal sebagai teks bersih (*clean text*) [1].

### a. Case Folding

*Case Folding* adalah langkah untuk mengonversi semua karakter dalam teks menjadi huruf kecil dan menghilangkan tanda baca serta angka. Proses ini hanya mempengaruhi huruf-huruf alfabet dari "a" hingga "z", sehingga karakter di luar rentang tersebut akan dihapus[2].

### b. Cleansing

*Cleansing* dilakukan untuk membersihkan tweet dari elemen yang tidak esensial seperti URL, hashtag '#', dan username '@'. URL diidentifikasi lewat format seperti http, https, atau [www](http://www). Tujuan dari proses ini adalah untuk mengurangi kebisingan dalam data [3].

- 1) Penghapusan Mention: Langkah ini mengeliminasi mention dalam teks, contohnya '@kitaBisa' akan dihilangkan.
- 2) Penghapusan Hashtag: Langkah ini menghilangkan kata-kata yang diawali dengan '#'. Misalnya, '#mudik' akan dihilangkan.
- 3) Penghapusan Link: Langkah ini menghilangkan tautan yang ada dalam teks. Sebagai contoh, <https://pbs.twimg.com/media/GKernYDaUAAPty8> akan dihilangkan.
- 4) Penghapusan Karakter Non-Alfanumerik: Langkah ini menghapus semua karakter selain huruf dan angka. Sebagai contoh, 'info mudik 2024!' akan diubah menjadi 'info mudik 2024'.
- 5) Penghapusan Spasi Berlebih: Langkah ini menghilangkan spasi ekstra dalam teks. Contoh, 'kabar mudik' akan diubah menjadi 'kabar mudik'.
- 6) Penghapusan Kata Pendek: Langkah ini menghapus kata-kata yang hanya terdiri dari satu atau dua huruf. Misalnya, 'a' dan 'di' akan dihapus dari teks.

### c. Mengubah Slang Word

Pengubahan kata-kata slang melibatkan konversi istilah gaul, singkatan, atau ungkapan informal menjadi bentuk yang lebih resmi. Sebagai contoh, 'utk' diubah menjadi 'untuk' dan 'yng' disesuaikan menjadi 'yang' [4].

### d. Menghapus Stop Word

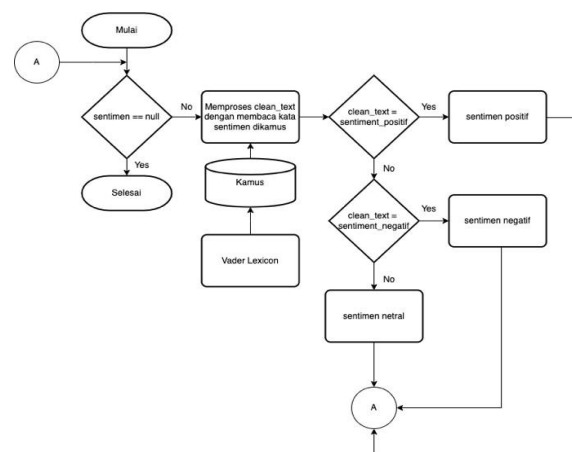
Proses eliminasi kata mencakup pengurangan kata-kata yang sering muncul namun memiliki sedikit kontribusi semantik dalam teks. Contoh kata-kata ini termasuk 'untuk', 'yang', dan 'apa' [5].

### e. Stemming

Penghapusan akhiran dalam kata dilakukan melalui proses yang dikenal sebagai stemming mengubah kata-kata berimbuhan ke bentuk dasarnya menggunakan perpustakaan Sastrawi. Sebagai contoh, 'mengurangi' diubah menjadi 'kurang', dan 'menaiki' menjadi 'naik' [6].

## 2.3. Labelling

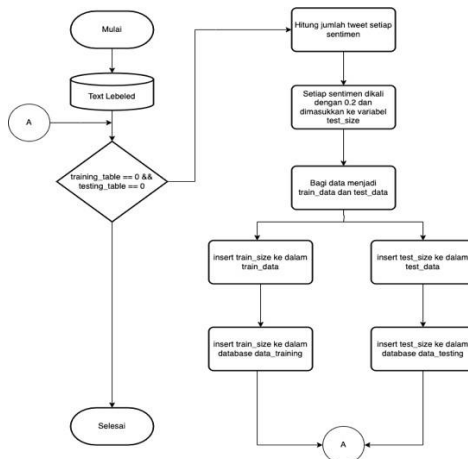
Pemberian label sentimen dilakukan dengan menerapkan kamus berisi kata-kata bernuansa positif dan negatif, yang telah disusun berdasarkan penelitian terdahulu dan tersedia di GitHub (<https://github.com/masdevid/ID-OpinionWords>). Proses anotasi ini memanfaatkan pendekatan yang berfokus pada kamus sentimen [7].



Gambar 1. Tahap Labelling

## 2.4. Split Data

Pada fase Pembagian Data, tweet yang telah dilabeli dibagi ke dalam dua grup: data latihan dan data uji. Secara umum, pembagian ini menetapkan 80% dari dataset sebagai data latihan dan 20% sebagai data uji [8]. Gambar 2 menunjukkan *flowchart* dari *Split* data.



**Gambar 2.** Tahap *Split* Data

## 2.5. TF-IDF

TF-IDF adalah teknik pembobotan yang populer dan efektif yang digunakan dalam pengambilan informasi dan dikenal memiliki performa yang baik dalam akurasi dan recall [9]. TF-IDF memberi bobot kepada kata berdasarkan frekuensinya dalam satu dokumen dan kelangkaannya di seluruh dokumen lain. Kata yang sering muncul dalam satu dokumen namun jarang di dokumen lainnya diberi bobot yang lebih tinggi, yang memperkuat peranannya dalam analisis, sedangkan kata yang sering muncul di banyak dokumen mendapat bobot lebih rendah. Berikut adalah persamaan untuk menghitung TF.

$$TF(d, t) = \frac{f(d, t)}{n(d)} \quad (1)$$

Pada persamaan ini,  $f(d, t)$  merepresentasikan frekuensi kemunculan kata tertentu dalam suatu dokumen, sementara  $n, d$  menunjukkan jumlah keseluruhan kata dalam dokumen tersebut. Rumus yang digunakan untuk menghitung IDF (Inverse Document Frequency) adalah sebagai berikut.

$$IDF(t) = \log \left( \frac{N}{df(t)} \right) \quad (2)$$

$N$  merepresentasikan jumlah keseluruhan dokumen dalam kumpulan, sementara  $df, t$  mengindikasikan jumlah dokumen yang mengandung kata  $t$  dalam kumpulan tersebut. Setelah nilai TF dan IDF diperoleh, langkah selanjutnya adalah menghitung nilai TF-IDF.

$$TF - IDF = TF(d, t) * IDF(t) \quad (3)$$

## 2.6. Confusion Matrix

*Confusion Matrix* adalah teknik yang umum digunakan untuk mengukur akurasi dalam *data mining*. Pada tabel 1 menunjukkan informasi tentang klasifikasi yang secara akurat diprediksi oleh sistem.

**Tabel 1.** *Confusion Matrix*

		Kelas Aktual		
		Positif	Netral	Negatif
Kelas Prediksi	Positif	TP	NtP	FP
	Netral	PNt	TNt	NNt
	Negatif	FN	NtN	TN

**Keterangan:**

**TP (True Positive)** = Dokumen yang sebenarnya positif dan benar diidentifikasi sebagai positif oleh model.

**FN (False Negative)** = Dokumen yang sebenarnya positif tapi diidentifikasi sebagai negatif oleh model.

**PNt (Positive Neutral)** = Dokumen yang sebenarnya positif tapi diidentifikasi sebagai netral oleh model.

**TN (True Negative)** = Dokumen yang sebenarnya negatif dan benar diidentifikasi sebagai negatif oleh model.

**FP (False Positive)** = Dokumen yang sebenarnya negatif tapi diidentifikasi sebagai positif oleh model.

**FNt (False Neutral)** = Dokumen yang sebenarnya negatif tapi diidentifikasi sebagai netral oleh model.

**NtP (Neutral Positive)** = Dokumen yang sebenarnya netral tapi diidentifikasi sebagai positif oleh model.

**NtN (Neutral Negative)** = Dokumen yang sebenarnya netral tapi diidentifikasi sebagai negatif oleh model.

**TNt (True Neutral)** = Dokumen yang sebenarnya netral dan benar diidentifikasi sebagai netral oleh model.

Tiga parameter yang perlu dihitung adalah akurasi, presisi, dan *recall*. Formula *Confusion Matrix* untuk mengkalkulasi akurasi, presisi, dan *recall* ditampilkan sebagai berikut.

$$accuracy = \frac{TP + TN}{TOTAL} \quad (4)$$

$$precision = \frac{TP}{TP + FP} \quad (5)$$

$$recall = \frac{TP}{TP + FN} \quad (6)$$

## 2.7. Naïve Bayes

*Naïve Bayes* berlandaskan pada *Teorema Bayes*, yang memberikan metode untuk menentukan peluang sebuah peristiwa berdasarkan peluang kondisional dari peristiwa itu sendiri dan peluang dari peristiwa yang berkaitan. [10] *Teorema Bayes* diungkapkan secara matematis melalui persamaan berikut:

$$\frac{P(H|E)}{P(H)} = \frac{P(E|H) \cdot P(H)}{P(E)} \quad (7)$$

**Deskripsi:**

E = Data yang kategorinya belum ditentukan.

H = Hipotesis kelas yang diusulkan untuk data E.

P(H|E) = Probabilitas terjadinya H, dengan pengetahuan mengenai E. P(E|H) = Probabilitas munculnya E, berdasarkan hipotesis H yang diketahui.

P(H) = Probabilitas sebelumnya dari H, dilihat dari peluang terjadinya H tanpa informasi tambahan. P(E) = Probabilitas dasar dari kejadian E.

Selanjutnya, probabilitas dasar untuk setiap kategori ditetapkan berdasarkan contoh dokumen. Dalam proses klasifikasi, kategori dokumen ditentukan dengan menganalisis istilah-istilah dalam dokumen yang akan diklasifikasikan.

## 2.8. Modelling

Selama tahap ini, terdapat delapan langkah utama yang harus dilakukan untuk mengembangkan model pelatihan. Proses-proses ini meliputi pemilihan dataset pelatihan, pengolahan token, pembuatan daftar kata, penghitungan vektor token, pengukuran Frekuensi Term, evaluasi Frekuensi Dokumen Terbalik, penggabungan Frekuensi Term dengan Frekuensi Dokumen Terbalik, serta penerapan Model Naïve Bayes.

## 2.9. Pengujian

Untuk mengukur sentimen terhadap layanan kepolisian selama periode mudik tahun 2024, dilakukan studi yang diimplementasikan melalui aplikasi web. Tahapan ini bertujuan untuk menghasilkan evaluasi performa dari klasifikasi data diilustrasikan menggunakan Matriks Kebingungan. Hasil dari klasifikasi ini akan dibagi ke dalam sembilan kategori, yaitu: positif akurat (true positive), positif keliru (false positive), positif netral (positive neutral), negatif akurat (true negative), negatif keliru (false negative), negatif netral (false neutral), netral akurat (neutral positive), netral keliru (neutral negative), dan netral netral (true neutral). Diharapkan, informasi ini akan memberikan pemahaman mengenai efektivitas metode yang diterapkan dalam analisis sentimen terkait layanan kepolisian selama periode mudik.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Tahap Pengumpulan Data

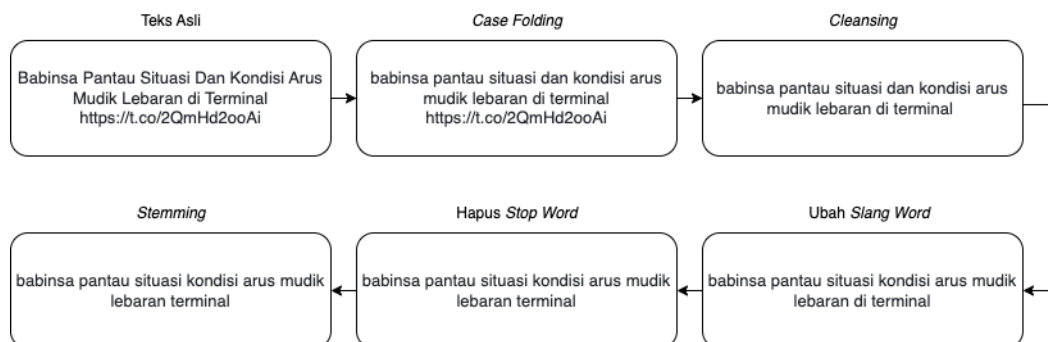
Informasi yang digunakan dalam penelitian ini didapatkan melalui proses *crawling* menggunakan aplikasi *Tweet Harvest*, dari rentang waktu 7 April 2024 hingga 14 April 2024, dan berhasil mengumpulkan sejumlah 1.176 data. Kata kunci yang digunakan dalam pengumpulan data ini adalah "info mudik" dan "kabar mudik". Selanjutnya, data ini akan diolah melalui tahap *Preprocessing*. Pada tabel 2 menunjukkan Data Penelitian.

**Tabel 2.** Data Penelitian

<i>Username</i>	<i>Full Text</i>	<i>Created_at</i>
balimus_tweet	LIVE REPORT : Situasi terkini arus mudik H-5 jelang Lebaran di pelabuhan Gilimanuk Jumat malam 5 April 2024. @infojembrana <a href="https://t.co/YJ6tv69yL2">https://t.co/YJ6tv69yL2</a>	Sat Apr 06 13:00:02 +0000 2024
aurorachavaaa	Kabar baik nih Resmi Dibuka! Mudik Gratis 2024 Pertamina Sediakan 3.500 Kuota Simak Cara Daftarnya. <a href="https://t.co/NYZQopKAd7">https://t.co/NYZQopKAd7</a>	Thu Mar 21 10:32:01 +0000 2024

#### 3.2. Tahap *Preprocessing*

Tahap *Preprocessing* dimulai setelah akumulasi satu atau lebih dataset dalam *database*. Langkah ini mencakup lima proses utama: *Case folding*, *Cleansing*, mengubah kata-kata *Slang*, menghilangkan *Stop Word*, dan *Stemming*. Pada gambar 3 menunjukkan alur dari tahap *Preprocessing*.



**Gambar 3.** Tahap *Preprocessing*

Dalam kajian ini, tahapan *Preprocessing* berhasil mengolah dan menghasilkan 1176 data *Tweet* menjadi teks yang lebih rapi, atau dikenal sebagai *clean text*. Setelah *Preprocessing*, data yang sudah diolah disimpan dalam *database* untuk digunakan pada tahap selanjutnya yaitu *Labelling*. Namun, beberapa data yang diproses tidak menghasilkan teks yang signifikan dan oleh karena itu tidak disertakan dalam *database*.

#### 3.3. Tahap *Labelling*

Setelah data bersih tersedia di *database* hasil dari tahap Pra-pemrosesan, proses pelabelan dimulai. Pada tahap ini, kamus sentimen digunakan untuk mengklasifikasikan sentimen dengan menghitung nilai skor sentimen.

**Tabel 3.** Tahap *Labelling*

<i>Clean Text</i>	Kata Positif	Kata Negatif	Kata Netral
security ngecek situasi biar warga aman mudik	Aman		
Jumlah	1	0	0

Tabel 3 menunjukkan tahap *labelling* yang dimana keberadaan satu kata positif dalam *clean text*, yang diidentifikasi berdasarkan frekuensi kata-kata positif yang muncul. Sebaliknya, tidak terdeteksi adanya kata negatif atau netral dalam *clean text*, menurut frekuensi kata-kata negatif dan netral yang muncul.

Setelah nilai skor diperoleh, langkah berikutnya adalah penentuan kelas sentimen berdasarkan aturan yang diuraikan dalam gambar 1.

```

if positive_count >
    negative_count: return
    'positif'
elif negative_count >
    positive_count: return 'negatif'
else:
    return 'netral'

```

Oleh karena itu, dapat disimpulkan bahwa Tweet yang berbunyi 'security ngecek situasi biar warga aman mudik' akan dikelompokkan ke dalam kelas positif. Dalam penelitian ini, total ada 945 Tweet yang sudah diberikan label.

### 3.4. Tahap Split Data

Pada tahap Pemisahan Data (*Split data*), tweet yang telah diberi label akan dibagi menjadi dua bagian: data training dan data testing. Proses ini menggunakan teknik *stratified sampling*.

Dalam penelitian ini, setiap tweet yang memiliki label positif, negatif, atau netral dihitung persentase masing-masing labelnya. Setelah persentase setiap label dihitung, data kemudian dibagi menjadi 80% untuk data training dan 20% untuk data testing berdasarkan persentase masing-masing label tersebut.

### 3.5. Tahap Modelling

Setelah tahap *Preprocessing*, *Labelling*, dan pemisahan data selesai, tahap selanjutnya adalah *Pemodelan*. Langkah ini bertujuan untuk menciptakan model pelatihan atau mengambil informasi dari data yang telah diproses. Proses ini melibatkan delapan langkah utama: pemilihan data pelatihan, tokenisasi, penyusunan kata, penghitungan vektor token, penghitungan Frekuensi Istilah, penghitungan Frekuensi Dokumen Terbalik, penggabungan Frekuensi Istilah dengan Frekuensi Dokumen Terbalik, dan pembangunan model *Naïve Bayes*. Hasil dari tahap *Pemodelan* ini adalah probabilitas *Likelihood* yang dapat ditunjukkan pada tabel 4.

**Tabel 4.** Tahap Probabilitas *Likelihood*

Kata	Probabilitas <i>Likelihood</i> Positif	Probabilitas <i>Likelihood</i> Positif	Probabilitas <i>Likelihood</i> Positif
kabar	0,130	0,074	0,058
gembira	0,074	0,043	0,058
diskon	0,074	0,043	0,058
akses	0,074	0,043	0,058
tol	0,074	0,043	0,058
gratis	0,074	0,043	0,058
mudik	0,074	0,043	0,058
lebaran	0,074	0,043	0,058
tips	0,074	0,043	0,058
jalur	0,074	0,043	0,058
darat	0,074	0,043	0,058
aman	0,074	0,043	0,058
nyaman	0,074	0,043	0,058
pantau	0,037	0,086	0,058

arus	0,037	0,086	0,117
hari	0,037	0,086	0,058
bahkan	0,037	0,086	0,058
duka	0,037	0,086	0,058
sedih	0,037	0,086	0,058
apa	0,037	0,086	0,058

Setelah menghitung probabilitas likelihood untuk setiap kata, langkah selanjutnya adalah mengubah data dari tabel tersebut menjadi file model (.model). Proses pemodelan ini memproduksi total 756 data tweet, yang terbagi menjadi 369 tweet berlabel positif, 126 tweet berlabel negatif, dan 261 tweet berlabel netral.

### 3.6. Tahap Klasifikasi Naïve Bayes

Tahap Klasifikasi dalam metode Naïve Bayes dijalankan berikutnya setelah pemodelan. Tujuan dari tahap ini adalah untuk mengklasifikasikan atau menetapkan label pada data uji menggunakan model yang telah terlatih. Sebagai persiapan, model yang telah dipilih digunakan untuk mengklasifikasikan data uji. Data uji yang digunakan adalah contoh yang ditunjukkan pada tabel 5.

**Tabel 5.** Tabel Data Uji

Data Uji	Teks Bersih	Label Sentimen
Uji 1	kabar baik bagi mau mudik terimakasih pertamina	Positif

Untuk menentukan kelas dari Data Uji, kita bisa menghitung probabilitas posterior untuk tiap kelas menggunakan metode *Naïve Bayes*.

- $$P(\text{positif}|\text{Uji 1}) = P(\text{positif}|\text{kabar}) * P(\text{positif}|\text{baik}) * P(\text{positif}|\text{bagi}) * P(\text{positif}|\text{mau}) * P(\text{positif}|\text{mudik}) * P(\text{positif}|\text{terimakasih}) * P(\text{positif}|\text{pertamina}) = 0,333 * 0,130 * 0,037 * 0,037 * 0,037 * 0,037 * 0,037 * 0,037 = 0,000000000632246$$
- $$P(\text{negatif}|\text{Uji 1}) = P(\text{negatif}|\text{kabar}) * P(\text{negatif}|\text{baik}) * P(\text{negatif}|\text{bagi}) * P(\text{negatif}|\text{mau}) * P(\text{negatif}|\text{mudik}) * P(\text{negatif}|\text{terimakasih}) * P(\text{negatif}|\text{pertamina}) = 0,333 * 0,072 * 0,37 * 0,037 * 0,037 * 0,037 * 0,037 = 0,000000000110702$$
- $$P(\text{netral}|\text{Uji 1}) = P(\text{netral}|\text{kabar}) * P(\text{netral}|\text{baik}) * P(\text{netral}|\text{bagi}) * P(\text{netral}|\text{mau}) * P(\text{netral}|\text{mudik}) * P(\text{netral}|\text{terimakasih}) * P(\text{netral}|\text{pertamina}) = 0,333 * 0,058 * 0,037 * 0,037 * 0,037 * 0,037 * 0,037 = 0,0000000004955443$$

**Tabel 6.** Tahap hasil prediksi *Naïve Bayes*

Data Uji	Teks Bersih	Label Sentimen	Probabilitas	Label Sentimen
Uji 1	kabar baik bagi mau mudik terimakasih pertamina	Positif	Positif = 0,000000000632246 Negatif = 0,000000000110702 Netral = 0,0000000004955443	Positif

Tabel 6 menunjukkan hasil prediksi menggunakan metode Naïve Bayes. Dari tabel tersebut, terlihat bahwa metode klasifikasi Naïve Bayes mampu melakukan prediksi yang akurat.

### 3.7. Hasil Pengujian

Prediksi telah dilakukan pada 1.689 data uji, yang direpresentasikan dalam bentuk Confusion Matrix. Berikut ini tabel 7 menunjukkan representasi dari Confusion Matrix tersebut.

**Tabel 7.** Confusion Matrix

		Kelas Aktual		
		Positif	Netral	Negatif
Kelas Prediksi	Positif	66	10	1
	Netral	4	38	17
	Negatif	20	5	28



Tabel 7 menunjukkan hasil pengujian yang dimana memungkinkan penggunaan persamaan (6), (7), dan (8) untuk menghitung nilai akurasi, presisi, dan recall. Hasil perhitungan untuk akurasi, presisi, dan *recall* dapat dilihat pada tabel 8.

**Tabel 8.** Hasil Akurasi, Presisi, dan *Recall*

Pengujian	
Akurasi	0.70 (70%)
Presisi	0.69 (69%)
<i>Recall</i>	0.72 (72%)

Hasil pengujian yang dilakukan dengan dataset yang telah dianalisis oleh ahli menunjukkan adanya peningkatan performa pada Algoritma Naïve Bayes. Setelah dianalisis, performa algoritma menunjukkan peningkatan akurasi menjadi 72%, presisi mencapai 73%, dan recall meningkat menjadi 75%. Evaluasi sentimen terhadap 1.176 data, yang mencakup data pelatihan dan pengujian, mengungkapkan bahwa sebagian besar warga X di Indonesia menunjukkan sentimen positif sebesar 51%, sentimen negatif sebesar 16%, dan sentimen netral sebesar 33%.

### 3. KESIMPULAN

Berdasarkan analisis dari 1.176 tweet yang dikumpulkan antara 7 April 2024 sampai 14 April 2024, terdapat 51% sentimen positif, 16% sentimen negatif, dan 33% sentimen netral yang dimana pada total dataset yang telah diberi label sebanyak 469 label positif, 157 negatif dan 319 netral dari total 945 data, menunjukkan bahwa sebagian besar masyarakat Indonesia memandang secara positif kinerja kepolisian dalam layanan mudik selama periode Lebaran tahun 2024. Peningkatan performa Algoritma Naïve Bayes setelah menggunakan dataset yang telah dianalisis oleh pakar memberikan bukti konkret tentang efektivitas analisis pakar dalam meningkatkan hasil algoritma dengan nilai akurasi 72%, presisi 73%, recall 75%, dan f1-score 71%. Saran untuk penelitian selanjutnya yaitu agar dilakukan perbaikan pada proses preprocessing data, pengembangan kamus sentimen yang lebih komprehensif, dan melakukan perbandingan dengan berbagai teknik pengujian lain untuk mengidentifikasi metode yang paling efektif. Langkah-langkah ini diharapkan dapat memperkuat kemampuan sistem dalam mengklasifikasikan sentimen secara lebih tepat dan efisien.

### DAFTAR PUSTAKA

- [1] D. Pramana, M. Afdal, M., and I. Permana, "Analisis Sentimen Terhadap Pemandangan Ibu Kota Negara Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbors," vol. 7, pp. 1306-1314, 2023.
- [2] A. Putri, and A. Muzakir, "Analisis Sentimen Cyberbullying KPOP Di Media Sosial Twitter Menggunakan Metode Naive Bayes," vol. 7, pp. 2541-0849, 2022.
- [3] N. S. Marga, A. R. Isnain, and D. Alita, "Sentimen Analisis Tentang Kebijakan Pemerintah Terhadap Kasus Corona Menggunakan Metode Naive Bayes," vol. 2, pp. 453-463, 2020.
- [4] G. N. Saputro Putri, D. Ispriyanti, and T. Widiarini, "Implementasi Algoritma Fuzzy C-Means Dan Fuzzy Possibilistics C-Means Untuk Klasterisasi Data Tweets Pada Akun Twitter Tokopedia," *Jurnal Gaussian*, vol. 11, no. 1, pp. 86-98, 2022.
- [5] I. B. Irawan, and O. Nurdiawan, "Naive Bayes Dan Wordcloud Untuk Analisis Sentimen Wisata Halal Pulau Lombok," *Infotech Journal*, vol. 9, no. 1, pp. 236-242, 2023.
- [6] C. H. Yutika, A., and S. Al Faraby, "Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, pp. 422-430, 2021.
- [7] P. G. Aryanti, and I. Santoso, "Analisis Sentimen Pada Twitter Terhadap Mobil Listrik Menggunakan Algoritma Naive Bayes," *Jurnal IKRAITH-Informatika*, vol. 7, no. 2, pp. 2580-4316, 2023.
- [8] I. A. Tanggraeni, and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI: Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 2, pp. 785-795, 2022.
- [9] M. R. Fais Sya' Bani, U. Enri, and T. N. Padilah, "Analisis Sentimen Terhadap Bakal Calon Presiden 2024 dengan Algoritma Naïve Bayes," *JURIKOM: Jurnal Riset Komputer*, vol. 9, no. 2, pp. 2407- 389X, 2022.
- [10] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 5, pp. 178-185, 2022.