

## ANALISIS SENTIMEN PENGAMBILALIHAN ASET PASCA DEMO DENGAN METODE NAÏVE BAYES

Abbashalom Radja<sup>1)</sup>, Karina Auliasari<sup>2)</sup>, Febriana Santi Wahyuni<sup>3)</sup>

<sup>1,2,3</sup> Teknik Informatika, Teknologi Industri, Institut Teknologi Nasional Malang  
Jln. Raya Karanglo Km.2 Malang, Jawa Timur, 65145

Co Responden Email: shalomradjaa@gmail.com

### Abstract

*This study discusses the issue of public sentiment toward the seizure of public officials' assets following the 2025 protest wave in Indonesia. This study aims to recognize and sort public opinions posted on social media platform X (formerly Twitter) sorted into three distinct attitudes classes positive, neutral, and negative using the Naïve Bayes algorithm. The research data was gathered using relevant keywords concerning asset forfeiture and looting, then underwent text preprocessing procedures such as tokenization, normalization, stopword filtering, and stemming. TF-IDF scoring was subsequently implemented before conducting sentiment analysis. The effectiveness of the model was evaluated using statistical measures such as accuracy, precision, recall, and F1-index indicators reveals that the Naïve Bayes approach efficiently sorts public opinions and recognizes the predominant sentiment trend. The study discovered that negative sentiment prevails, reflecting significant public criticism of post-demonstration asset seizures. These outcomes suggest that Naïve Bayes represents a practical solution for analyzing public perspectives on social and political topics in Indonesia.*

### Article history

Received 11 Dec 2025

Revised 30 Dec 2025

Accepted 20 Jan 2026

Available online 31 Jan 2026

### Keywords

Sentiment Analysis,  
Naïve Bayes,  
Public Opinion,  
Asset Seizure,  
Social Media,

### Abstrak

Penelitian ini membahas isu sentimen publik terhadap pengambilalihan aset milik pejabat publik setelah gelombang protes tahun 2025 di Indonesia. Penelitian ini bertujuan untuk mengenali serta mengelompokkan opini masyarakat yang diunggah di media sosial X (Twitter) ke kategori positif, negatif, netral menggunakan algoritma *Naïve Bayes*. Dataset dikumpulkan melalui kata kunci terkait isu penjarahan dan pengambilalihan aset, kemudian diproses melalui tahapan tokenisasi, normalisasi, penghapusan *stopword*, dan *stemming*. Pembobotan *TF-IDF* diterapkan sebelum dilakukan klasifikasi sentimen. Hasil pengujian dievaluasi melalui ukuran kinerja menggunakan akurasi, presisi, recall, *F1-score* menunjukkan algoritma *Naïve Bayes* mampu mengklasifikasikan opini dengan baik dan mengidentifikasi sentimen dominan. Analisis menunjukkan bahwa sentimen negatif lebih mendominasi, mencerminkan ketidaksetujuan masyarakat terhadap pengambilalihan aset pasca protes. Temuan ini terbukti bahwa metode *Naïve Bayes* efektif digunakan untuk memahami persepsi publik terhadap isu sosial dan politik di Indonesia.

### Riwayat

Diterima 11 Des 2025

Revisi 30 Des 2025

Disetujui 20 Jan 2026

Terbit online 31 Jan 2026

### Kata Kunci

Analisis Sentimen,  
Naïve Bayes,  
Opini Publik,  
Pengambilalihan aset,  
Sosial Media,

## PENDAHULUAN

Gelombang demonstrasi yang terjadi di Indonesia pada tahun 2025 memunculkan dampak sosial yang luas dan berkelanjutan. Setelah aksi protes mereda, sebagian massa melakukan pengambilalihan atau penjarahan aset milik pejabat publik serta fasilitas umum. Fenomena ini memicu kekhawatiran sosial,

kerugian ekonomi, dan menurunnya rasa aman masyarakat terhadap perlindungan aset pribadi maupun publik. Kondisi tersebut mencerminkan pergeseran pola protes di Indonesia, di mana gerakan sosial tidak hanya bertujuan menyuarakan aspirasi politik, tetapi juga dapat berkembang menjadi tindakan kolektif yang berpotensi mengancam stabilitas

sosial dan kepercayaan publik terhadap pemerintah.

Dalam konteks ini, penting untuk memahami bagaimana persepsi masyarakat terbentuk terhadap isu pengambilalihan aset pasca demonstrasi. Pendapat masyarakat menjadi peran penting dalam membangun legitimasi sosial dan menentukan arah kebijakan pemerintahan. Dengan menggunakan metode analisis sentimen, perspektif masyarakat dapat diklasifikasikan secara terstruktur ke kategori positif, negatif, netral. Pendekatan analisis seperti ini mampu menggambarkan sejauh mana masyarakat mendukung atau menentang isu-isu sosial politik tertentu, serta dapat dijadikan landasan bagi para pengambil keputusan untuk menyusun strategi komunikasi publik yang lebih tepat sasaran.

Studi ini menerapkan algoritma *Naïve Bayes*, yaitu sebuah teknik klasifikasi yang berdasarkan pada probabilitas dengan karakteristik sederhana tetapi efektif untuk mengolah data teks dalam jumlah besar. Beberapa riset sebelumnya telah membuktikan kemampuan pendekatan ini. Amini dan Setiawan (2024) berhasil menerapkan *Naïve Bayes* dalam analisis opini masyarakat terhadap calon wakil presiden 2024 dengan akurasi 82,19%. Penelitian lain oleh Mantika et al. (2024) menunjukkan bahwa *Naïve Bayes* tetap kompetitif dibandingkan metode lain seperti *Logistic Regression*. Sementara itu, Gunawan et al. (2024) menunjukkan efektivitas metode ini dalam mengklasifikasikan opini politik di media sosial dengan akurasi 85%.

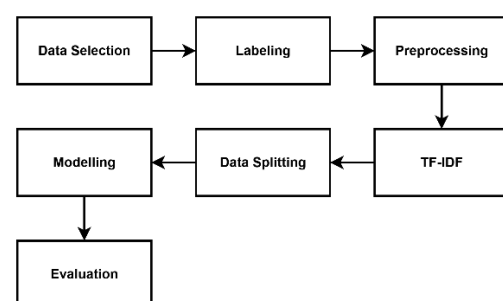
Berdasarkan temuan-temuan tersebut, penelitian dilakukan guna menganalisis sentimen masyarakat di media sosial X (*Twitter*) terhadap isu pengambilalihan aset pasca gelombang protes 2025. Melalui pendekatan *text mining* dan pembobotan *TF-IDF*, data teks publik diolah menjadi informasi terstruktur yang dapat digunakan untuk mengukur persepsi sosial secara kuantitatif. Tujuan penelitian ini adalah untuk mengidentifikasi sentimen dominan masyarakat, mengukur performa algoritma *Naïve Bayes* dalam klasifikasi opini, serta memberikan gambaran empiris mengenai arah opini publik terhadap fenomena pengambilalihan aset. Hasil temuan dari studi

ini diharapkan dapat memberikan kontribusi pada pengembangan analisis sentimen di ranah sosial dan politik dan menjadi referensi dalam pemanfaatan teknologi pemrosesan bahasa alami (*Natural Language Processing*) di Indonesia

## METODE PENELITIAN

Dalam pelaksanaannya, penelitian ini mengadopsi pendekatan kuantitatif yang dianalisis menggunakan metode analisis sentimen yang berbasis pada pembelajaran mesin. Objek penelitian adalah opini publik di media sosial X yang berkaitan dengan isu pengambilalihan aset pejabat publik pasca gelombang protes tahun 2025. Proses pengumpulan data dibuat melalui teknik *web crawling* dengan menggunakan kata kunci sesuai, kemudian data tersebut dikelompokkan ke tiga kategori sentiment (positif, netral, negatif). Rangkaian kegiatan penelitian mencakup tahap pengumpulan data, prapemrosesan teks, penerapan pembobotan *TF-IDF*, proses klasifikasi dengan algoritma *Naïve Bayes*, serta penilaian performa model melalui *confusion matrix* yang menggunakan parameter akurasi, presisi, *recall*, dan *F1-score*.

Kerangka metodologi penelitian ini terbagi menjadi tujuh fase pokok, seperti yang terlihat pada Gambar 1. Fase-fase tersebut mencakup seluruh proses mulai dari pengumpulan data sampai dengan penilaian kinerja model klasifikasi sentimen.



Gambar 1. Tahapan metode uji data

Alur dari aplikasi pertama dilakukannya pengumpulan data adalah mengumpulkan data dari *platform* media sosial dengan melakukan pencarian berdasarkan kata kunci yang spesifik. Klasifikasi: Informasi yang diambil dari media sosial dikategorikan berdasarkan sentimen menjadi kelompok positif, negatif, atau netral. Prapemrosesan: Pada tahap ini, teks dibersihkan, dinormalisasi, dipecah

menjadi bagian-bagian kecil (tokenisasi), dan dibersihkan dari kata-kata umum (*stop words*) serta dilakukan *stemming* dengan menggunakan perpustakaan Sastrawi. Ekstraksi Fitur: Dengan metode TF-IDF (*Term Frequency-Inverse Document Frequency*), teks diubah menjadi bentuk numerik untuk analisis lebih mendalam. Pembagian Dataset: Data dibagi menjadi subset yang digunakan untuk pelatihan dan pengujian model. Pembentukan Model: Algoritma *Multinomial Naive Bayes* dilatih menggunakan data pelatihan untuk mendeteksi pola sentimen. Evaluasi: Model diuji dengan data pengujian. Proses evaluasi ini melibatkan penghitungan akurasi, presisi, *recall*, dan *F1-score* untuk menilai seberapa baik model berfungsi. Aplikasi ini mengotomatisasi seluruh proses ini guna melakukan analisis sentimen dengan cara yang efisien dan sistematis.

### 1. Data Selection

Sumber informasi yang digunakan yakni data utama diambil dari media sosial X (*Twitter*) dengan Teknik *Crawling*. *Crawling* dataset didasarkan pada kata kunci seperti “pengambilalihan aset pejabat”, “penjarahan properti pejabat”, “protes 2025”, dan “pengambilalihan aset pasca demonstrasi”. Tweet yang dikumpulkan adalah unggahan berbahasa Indonesia dengan rentang waktu sebulan. Setelah proses pengumpulan, data disimpan dalam format CSV dengan atribut *tweet*, tanggal, *username*, dan label.

### 2. Data Labeling

Labeling dilakukan secara manual pada data training. *Tweet* dikategorikan menjadi tiga kelas sentimen, yaitu: positif (mendukung), netral (informatif), dan negatif (menolak atau berkonotasi kecaman).

### 3. Preprocessing

Tahapan ini mencakup beberapa tahapan : (a) *Cleaning* karakter khusus seperti *mention*, *hashtag*, dan *URL*; (b) *Case Folding*; (c) *Tokenizing*; (d) *Stopword Removal*; (e) *Stemming*; dan (f) *lemmatization*. Berikut tahap-tahap *preprocessing*:

#### a. Cleaning

Pada tahap ini bertujuan mengubah data *raw* atau mentah (hasil *scraping*) menjadi teks bersih dari elemen yang tidak relevan.

#### b. Case Folding

Tahap ini merubah huruf besar pada teks menjadi huruf kecil, membantu menghindari perbedaan antar huruf besar kecil saat menganalisis.

#### c. Tokenizing

Setelah *Case Folding* untuk memecah teks panjang menjadi perkata (per kata atau per frasa) agar mudah dilanjutkan ke proses selanjutnya.

#### d. Stopword Removal

Langkah selanjutnya berfungsi guna menghilangkan kata-kata di daftar *stopwords*. *Stopwords* merupakan kata-kata biasanya tidak diperhatikan pada analisis teks, seperti "dan", "di", "itu". Dengan membuang *stopwords*, model dapat lebih memusatkan perhatian ke kata-kata lebih penting.

#### e. Stemming

Tahapan merubah setiap kata kemudian dikembalikan bentuk kata dasarnya (*stemming*). Ini serupa dengan lematisasi, tetapi lebih agresif dalam menghapus akhiran kata.

#### f. Lemmatization

Tahapan ini merubah sebuah teks kembali ke kata dasarnya (lematisasi). Guna mempermudah menyaring kata dasarnya.

### 4. Feature Extraction

Tahap ini melibatkan penggunaan metode TF-IDF (*Term Frequency-Inverse Document Frequency*), merupakan teknik memberikan nilai atau bobot pada kata-kata untuk mengubah bentuk teks menjadi data numerik. Bobot yang diperoleh ini kemudian dimanfaatkan untuk klasifikasi data (Natasuwarna, 2019). *TF-IDF* terdiri dari dua komponen: frekuensi istilah (*TF*) dan frekuensi dokumen terbalik (*IDF*). *TF* mengukur frekuensi kemunculan suatu kata pada suatu dokumen; makin tinggi tingkat kemunculan istilah tersebut, maka bobot *TF* yang dihasilkan pun meningkat. Komponen *IDF* menghitung seberapa sering sebuah kata muncul di keseluruhan korpus teks, dan nilainya akan berbanding terbalik dengan jumlah dokumen yang mengandung kata itu. Jika sebuah istilah tersebut muncul di sejumlah

besar dokumen, nilai IDF-nya akan semakin mendekati 0 (Winarti et. al., 2017).

## 5. Data Splitting

Setelah melakukan *preprocessing*, data dibagi menjadi dua, 80% data *training*, 20% data *testing*. Pemisahan dilakukan secara acak dengan parameter *random\_state* yakni 42 agar hasil dapat direproduksi.

## 6. Modeling

Setelah melakukan data *splitting*, Model digunakan pada penelitian ini adalah *Multinomial Naïve Bayes*. Berdasarkan prinsip teorema *Bayes* dengan asumsi independensi antar fitur. Model dibangun menggunakan data *training*, dan dilakukan pelatihan untuk menghasilkan probabilitas tiap kelas sentimen. Dengan menggunakan rumus teorema *Bayes* berikut :

$$P(C|X) = \frac{P(C|X) \times P(C)}{P(X)} \quad (1)$$

Pada persamaan (1) diteorema bayes digunakan untuk menghitung probabilitas dalam suatu teks ke kelas tertentu dari kata-kata yang muncul didalam persamaan tersebut  $P(C|X)$  merupakan persamaan untuk menghitung probabilitas teks ke kelas C setelah diketahui kata-kata ataupun fitur yang muncul dipersamaan tersebut,  $P(X|C)$  adalah probabilitas kata-kata kemunculan dalam kelas tertentu seperti C,  $P(C)$  dari kelas tersebut menunjukkan probabilitas prior ataupun awal dari kelas c sebelum mempertimbangkan fitur X, dan  $p(x)$  merupakan probabilitas kemunculan kata-kata X dengan keseluruhan dari kelas tersebut. Dengan menggunakan persamaan ini, sistem memperbarui keyakinan dari suatu kelas dengan informasi yang ditemukan dalam suatu teks, sehingga memungkinkan ditentukannya seperti positif, negatif, dan netral secara probabilistik.

## 7. Implementation

Dalam fase implementasi, penelitian ini memanfaatkan beberapa alat pengembangan yang bekerja sama untuk mendukung proses analisis sentimen. *Python* dipilih sebagai bahasa pemrograman utama karena menawarkan pustaka yang cukup lengkap untuk pengolahan teks, termasuk untuk penerapan metode *Multinomial Naive Bayes*.

Aplikasi ini dikembangkan menggunakan *Flask*, yang dinilai lebih ringan dan mudah disesuaikan dengan kebutuhan sistem berbasis web. Untuk normalisasi teks berbahasa Indonesia, penelitian ini menggunakan *Sastrawi* sebagai alat untuk *stemming* dan penghapusan *stopword*. Sementara itu, *MySQL* digunakan sebagai media penyimpanan data karena kestabilannya dan kemudahan integrasinya dengan *Flask*. Kombinasi perangkat ini diterapkan agar alur pemrosesan data dan pengembangan aplikasi bisa berjalan lebih terstruktur dan sesuai dengan tujuan penelitian.

## 8. Evaluation

Tahapan evaluasi bertujuan untuk menilai kinerja dari klasifikasi yang diterapkan. Data diuraikan, prediksi dilakukan ke dalam matriks kebingungan. Perhitungan akurasi, presisi, kemudian *recall* digunakan untuk menilai performa dari model. Matriks kebingungan merupakan tabel yang berfungsi sebagai indikator untuk mengetahui jumlah data uji terklasifikasi dengan benar atau salah dari model yang telah diimplementasikan. Matriks kebingungan terdiri dari beberapa elemen, yakni *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). *TP* merupakan hasil prediksi positif yang memang benar-benar memiliki nilai positif, sementara *TN* adalah hasil prediksi negatif yang sesungguhnya juga bernilai negatif. Di sisi lain, *FP* menunjukkan hasil prediksi positif meski nilai sebenarnya adalah negatif, dan *FN* menggambarkan hasil prediksi negatif meskipun nilai aktualnya adalah positif. Metode pada *confusion matrix* tersebut digunakan untuk menghitung akurasi, presisi, *recall*, dan *F1-score* dengan rumus:

$$\text{Akurasi} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (2)$$

$$\text{Presisi} = \frac{TP}{(TP + FP)} \quad (3)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (4)$$

$$\text{F1 Score} = 2 \times \frac{(\text{Presisi} \times \text{Recall})}{(\text{Presisi} + \text{Recall})} \quad (5)$$

Pada persamaan (2), (3), (4), dan (5), beberapa metrik evaluasi digunakan untuk menilai seberapa baik model klasifikasi bekerja,

terutama dalam mengevaluasi prediksi berdasarkan data sebenarnya. Persamaan (2) menjelaskan akurasi, yang dihitung menggunakan rumus

$$\text{Akurasi} = (TP + TN) / (TP + TN + FP + FN),$$

di mana *TP* (Positif Benar) merujuk pada jumlah data yang diklasifikasikan dengan tepat sebagai positif, *TN* (Negatif Benar) adalah data yang diklasifikasikan dengan benar sebagai negatif, *FP* (Positif Salah) adalah data yang salah diklasifikasikan sebagai positif, dan *FN* (Negatif Salah) adalah data yang salah diklasifikasikan sebagai negatif. Persamaan (3) menjelaskan presisi sebagai:

$$\text{Presisi} = TP / (TP + FP),$$

yang menilai keakuratan model dalam memprediksi data positif di antara semua data yang diprediksi positif. Persamaan (4) mendefinisikan recall sebagai:

$$\text{Recall} = TP / (TP + FN),$$

yang menyoroti keyakinan model untuk mengidentifikasi keseluruhan data positif yang ada. Sementara itu, persamaan (5) adalah *F1-Score*, yang dapat dihitung menggunakan formula  $F1\ Score = 2 \times (\text{Presisi} \times \text{Recall}) / (\text{Presisi} + \text{Recall})$ , yang berguna untuk mengevaluasi keseimbangan antara presisi dan *recall*. Keempat metrik ini digunakan secara bersamaan untuk memberikan gambaran yang komprehensif tentang kinerja model, meliputi aspek akurasi, presisi, serta bagaimana model menjaga keseimbangan dalam melaksanakan tugas klasifikasi.

## HASIL DAN PEMBAHASAN

Serangkaian pengujian dilakukan pada dataset yang didapatkan dari hasil *scraping* jejaring sosial *Twitter* atau *X* yang mencakup 2.218 tweet dalam bahasa Indonesia mengenai isu pengambilalihan aset pejabat publik setelah protes besar di tahun 2025. Dataset ini telah melalui proses anotasi manual dan dibagi menjadi tiga kategori sentimen (positif, netral, negatif). Menurut hasil anotasi tersebut, sentimen negatif mendominasi dengan proporsi mencapai 52,3%, diikuti oleh sentimen netral 34,9%, sentimen positif hanya 12,8%. Distribusi ini mengindikasikan bahwa

opini publik cenderung memiliki sentimen negatif, yang mencerminkan ketidaksetujuan masyarakat terhadap tindakan mengambil alih aset pejabat publik setelah gelombang protes tahun 2025.

Tabel 1. Contoh pelabelan berdasarkan *tweet*

Label	Tweet
Netral	Polisi Tangkap Pelaku Penjarahan Rumah Menkeu Sri Mulyani <a href="https://t.co/hSz0tI9xVD">https://t.co/hSz0tI9xVD</a>
Negatif	@ARSIPAJA Sepertinya ibu ini tidak sadar bahwa dirinya selama ini telah lakukan penjarahan secara masal dan lebih besar besaran pada rakyat ini untuk biayai kelakuan konco konconya di pemerintahan dan parlemen.
positif	Mari Jaga Kondusifitas Tolak Anarkis Penjarahan dan Mari kawal aspirasi dengan cara Damai dan Bermartabat <a href="https://t.co/hq6kMg3X5Y">https://t.co/hq6kMg3X5Y</a>

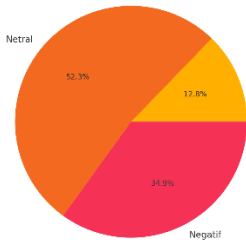
### 1. Analisis Data

Dataset penelitian dianalisis menggunakan visualisasi *Pie Chart* dan *Word Cloud* agar membantu mempermudah pemahaman.

### 2. Visualisasi Data

Untuk memahami kata yang paling sering muncul di setiap kategori sentimen, dilakukan visualisasi menggunakan *Word Cloud*. Gambar berikut menampilkan hasil visualisasi untuk kategori netral. Model *Naïve Bayes* diuji pada dataset berjumlah 444 *tweet*. Distribusi data menunjukkan bahwa sentimen negatif mendominasi sebesar 52,3%, netral 34,9%, dan positif 12,8%. Evaluasi model menghasilkan akurasi 73,0%, dengan presisi sebesar 72,6%, *recall* sebesar 73,0%, dan *F1-score* sebesar 72,1%. Hasil ini menunjukkan bahwa model mampu mengklasifikasikan opini publik dengan performa yang stabil dan akurat.

Distribusi Sentimen Publik terhadap Isu Pengambilalihan Aset



Gambar 2. Pie chart kategori sentimen

Di tahap visualisasi Awan Kata tersebut, dapat menampilkan kata-kata sering munculnya dalam dataset tersebut.



Gambar 3. Awan kata negatif



Gambar 4. Awan kata netral



Gambar 5. Awan kata positif

Pada Gambar 3,4 dan 5 bisa dilihat bahwa kata paling sering muncul adalah 'penjarahan', 'rumah', 'fasilitas', 'umum', dan 'aksi'. Kata-kata ini menunjukkan bahwa wacana publik cenderung berfokus pada peristiwa pengambilalihan aset serta keterlibatan masyarakat dan pejabat publik. Selain itu, kata seperti 'tegas' dan 'presiden' menandakan munculnya harapan agar pemerintah bersikap tegas dalam menindak aksi tersebut.

### 3. Hasil Klasifikasi Naïve Bayes

Dalam mengklasifikasikan sentimen ini, algoritma yang digunakan yaitu *Multinomial Naïve Bayes* dengan pembobotan *TF-IDF*. Evaluasi model dilakukan dengan data uji sebanyak 20% dari total dataset. Dimana hasil pengujian tersebut dalam tabel 2.

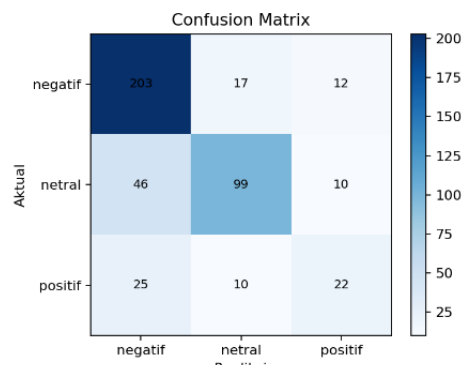
Tabel 2. Laporan hasil klasifikasi sentimen

Sentimen	Negatif	Netral	Positif	Akurasi
Presisi	0.741	0.786	0.500	0.730
Recall	0.875	0.639	0.386	0.730
F1-Score	0.802	0.705	0.436	0.730
Support	232	155	57	0.730

Nilai akurasi keseluruhan model mencapai 73,0%, dengan presisi sebesar 0,726%, recall sebesar 0,730%, dan F1-score sebesar 0,721%. Performa tersebut memperlihatkan bahwa metode *Naïve Bayes* berhasil mengelompokkan opini masyarakat dengan tingkat akurasi yang memadai, terutama dalam mengidentifikasi sentimen negatif.

### 4. Confusion Matrix

Visualisasi pada *confusion matrix* dalam Gambar 6 menunjukkan distribusi hasil prediksi model terhadap data aktual.

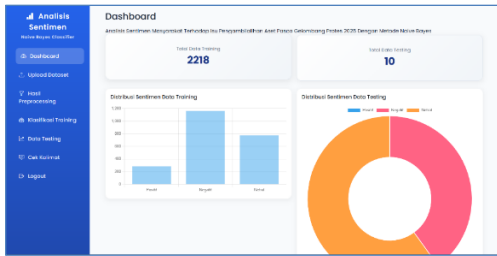


Gambar 6. Hasil Gambar Confusion matrix

Pada Gambar 6, model memiliki performa terbaik dalam mendeteksi sentimen negatif, dengan 203 prediksi benar dari total 232 data aktual. Namun, model masih kesulitan mengenali kelas positif, hanya 22 data dari 57 data, sebab data untuk kelas positif dalam pelatihan. Kelas netral pun menunjukkan 99 *tweet* dikenali dengan benar dari 155 *tweet*, menandakan bahwa 30% opini netral berakhir terklasifikasi sebagai negatif.

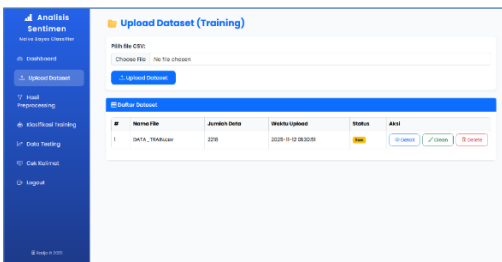
### 5. Tampilan Aplikasi

Terdapat berbagai fitur pada aplikasi yang mana tiap fitur tersebut memiliki fungsinya masing-masing.



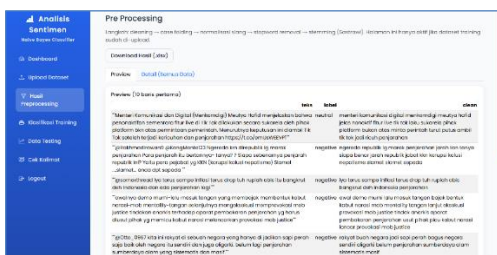
Gambar 7. Tampilan dashboard

Halaman ini menampilkan hasil implementasi dari distribusi data antar sentimen. Berikut tampilan dashboard yang ada pada gambar 7.



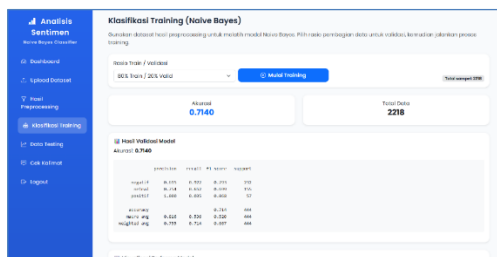
Gambar 8. Tampilan upload dataset

Halaman ini menampilkan tempat untuk digunakan mengunggah serta mengelola dataset. Berikut tampilan upload dataset yang ada pada gambar 8.



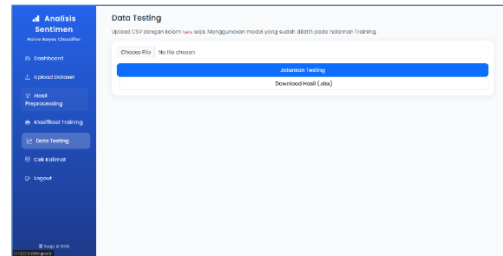
Gambar 9. Tampilan hasil preprocessing

Halaman ini menampilkan hasil dari data yang sudah dibersihkan dan hasil pemrosesan teks. Berikut tampilan hasil preprocessing pada gambar 9.



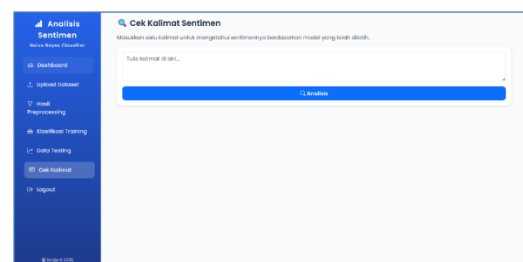
Gambar 10. Tampilan klasifikasi training

Halaman ini menampilkan hasil dari pelatihan model serta evaluasinya. Berikut tampilan hasil klasifikasi training pada gambar 10.



Gambar 11. Tampilan data testing

Halaman ini menampilkan penggunaan dataset yang baru untuk dilakukan pengujian model. Berikut tampilan data testing pada gambar 11.

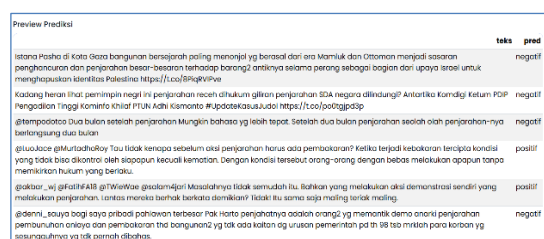


Gambar 12. Tampilan cek kalimat

Halaman ini menampilkan untuk menguji sentimen dari input satu kalimat saja. Berikut tampilan cek kalimat pada gambar 12.

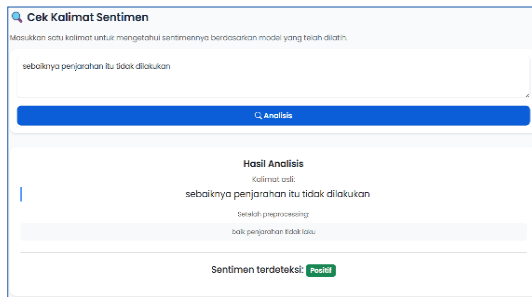
### 6. Validasi Hasil Pengujian

Validasi hasil dilakukan untuk memastikan bahwa aplikasi mampu menjalankan proses klasifikasi sentimen secara konsisten pada data yang tidak termasuk dalam data pelatihan. Pengujian pertama dilakukan melalui fitur Data Testing dengan menggunakan dataset baru. Aplikasi berhasil memproses data tersebut dan menghasilkan prediksi sentimen yang sesuai dengan pola distribusi utama pada penelitian ini. Hasil ini membuktikan bahwa model memiliki kemampuan generalisasi yang baik terhadap data eksternal.

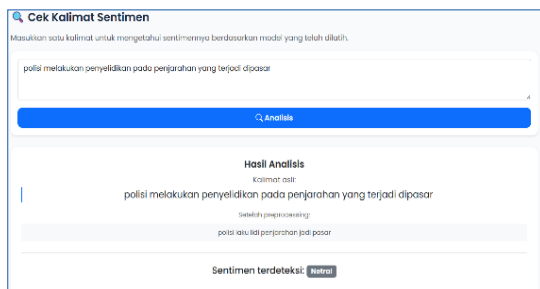


Gambar 13. Data testing dengan data baru

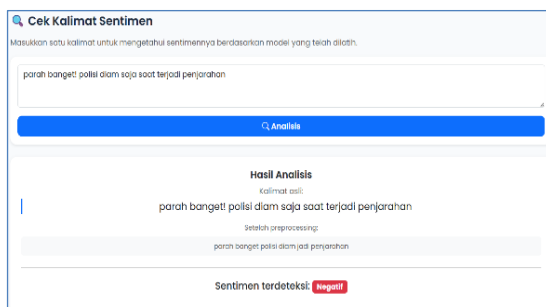
Selanjutnya, pengujian dilakukan menggunakan fitur Cek Kalimat untuk menganalisis satu kalimat secara langsung. Prediksi sentimen yang dihasilkan sistem sesuai dengan konteks kalimat yang diberikan, sehingga menunjukkan bahwa model bekerja secara operasional dan mampu memberikan hasil *real-time* dengan akurasi yang dapat dipertanggungjawabkan.



Gambar 14. Uji *input* kalimat positif



Gambar 15. Uji *input* kalimat netral



Gambar 16. Uji *input* kalimat negatif

Kedua pengujian ini membuktikan bahwa aplikasi tidak hanya valid secara statistik melalui metrik evaluasi, tetapi juga dapat digunakan secara fungsional untuk menganalisis data baru maupun masukan individu secara langsung.

## KESIMPULAN

Menilik hasil analisis, tampak jelas bahwa Sentimen negatif menguasai percakapan publik, menyoroti kecenderungan masyarakat yang tak segan mengkritik

tindakan pengambilalihan aset oleh pejabat publik.

Model ini mencapai akurasi 73,0%, menegaskan bahwa gabungan tahapan preprocessing, vektor *TF-IDF*, serta algoritma *Naïve Bayes* terbukti efektif dalam menelaah dan memetakan opini publik.

Kinerja paling tinggi muncul di kelas negatif dengan *recall* sebesar 87.5 %, sementara kelas positif menunjukkan kinerja terendah karena datanya yang terbatas.

Temuan ini sejalan dengan penelitian Amini & Setiawan (2024) serta Gunawan et al. (2024), yang memang mengindikasikan bahwa *Naïve Bayes* bekerja secara optimal dalam analisis sentimen berbahasa Indonesia. Studi ini mengungkap bahwa algoritma *Naïve Bayes* terbukti cukup ampuh untuk menelaah pendapat masyarakat tentang pengambilalihan aset pasca gelombang protes 2025.

Di antara spektrum sentimen yang terdeteksi, nuansa negatif muncul paling menonjol, menandakan ketidaksetujuan publik yang meluas terhadap fenomena tersebut. Dengan demikian, model ini dapat dijadikan pijakan bagi penelitian lanjutan yang menggali opini publik melalui media sosial.

## REFERENSI

Amini, T., & Setiawan, K. (2024). Application of the Naïve Bayes Algorithm in Twitter Sentiment Analysis of 2024 Vice Presidential Candidate Gibran Rakabuming Raka Using Rapidminer. *International Journal of Software Engineering and Computer Science (IJSECS)*, 4(1), 234–246.

Mantika, A., Triayudi, A., & Aldisa, R. (2024). Sentiment Analysis on Twitter Using Naïve Bayes and Logistic Regression for the 2024 Presidential Election. *SaNa: Journal of Blockchain, NFTs and Metaverse Technology*, 2(1), 44–55.

Gunawan, Y., Purnama, I., & Rohani, R. (2024). *Sentiment Analysis of Twitter Towards the 2024 Indonesian Presidential Candidates Using the Naïve Bayes Algorithm*. *International Journal of Science, Technology & Management*, 5(4), 953–961.

- Spierenburg, L., van Cranenburgh, S., & Cats, O. (2025). *Studying Riots Through the Lens of Social Media*. *Journal of Big Data*, 12(182), 1–15.
- Sheikhi, R. A., Javanbakhtian, R., & Heidari, M. (2025). *Looting and Antisocial Behavior After Disasters: A Systematic Review*. *BMC Public Health*, 25(309), 1–12.
- Azhari, A., & Hadi, M. (2023). *Multinomial Naïve Bayes for Sentiment Analysis of Indonesian Local Government Performance*. *Signal and Image Processing Letters*, 3(2), 1–10.
- Khairunnisa, K., Dewi, S. K., Rahmawati, D. D., & Sari, A. P. (2024). *Analisis Sentimen Komentar pada Postingan Instagram Akun “StandWithUs” Menggunakan Klasifikasi Naïve Bayes*. *Jurnal Ilmiah Informatika*, 12(2), 191–199.
- Khoerunnisa, S., Shiddieq, D. F., & Nurhayati, D. (2025). *Penerapan Algoritma Naive Bayes dengan Teknik TF-IDF dan Cross Validation untuk Analisis Sentimen terhadap Starlink*. *Malcom: Indonesian Journal of Machine Learning and Computer Science*, 5(2), 566–577.
- Malik, T., Hanif, N., Tahir, A., Abbas, S., Hanif, M. S., Tariq, F., Ansari, S., Abbasi, Q. H., & Imran, M. A. (2023). *Crowd Control, Planning, and Prediction Using Sentiment Analysis: An Alert System for City Authorities*. *Applied Sciences*, 13(3), 1592.
- Herdiyani, T. C., & Zailani, A. U. (2022). *Sentiment Analysis Terkait Pemindahan Ibu Kota Indonesia Menggunakan Metode Random Forest Berdasarkan Tweet Warga Negara Indonesia*. *Jurnal Teknologi Sistem Informasi*, 3(2), 154–165.
- Permana, A. A., & Putra, P. P. (2023). *Sentiment Analysis of Public Opinion of Citayam Fashion Week on Social Media Twitter Using Naïve Bayes Classifier*. *Jurnal Informatika Dharmakarya*, 7(1), 112–116. (Sitasi dari Jurnal Dharmakarya)
- Suryana, R., & Dini, F. (2023). *Analisis Sentimen Masyarakat terhadap Program Digitalisasi Pemerintahan Menggunakan Algoritma Naïve Bayes*. *Jurnal Dharmakarya Informatika*, 8(2), 215–224. (Sitasi dari Jurnal Dharmakarya)
- Natasuwarna, A. P. (2019). *Analisis Sentimen Keputusan Pemindahan Ibukota Negara Menggunakan Klasifikasi Naive Bayes*. *Seminar Nasional Sistem Informasi & Teknologi Informasi (SENSITIF)*, 355–363.
- Winarti, T., Kerami, J., & Arif, S. (2017). *Determining Term on Text Document Clustering Using Algorithm of Enhanced Confix Stripping Stemming*. *International Journal of Computer Applications*, 157(9), 3156–3164.
- Rahmawati, L. (2022). *Analisis Sentimen terhadap Kebijakan Pemerintah Menggunakan Naïve Bayes Classifier (Studi Kasus: Penghapusan Subsidi BBM)*. (Unpublished undergraduate thesis). Universitas Negeri Malang, Malang, Indonesia.