

## PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN OPINI MASYARAKAT PADA DATA TWITTER

Al Hajju Arafah<sup>1\*</sup>, Rizky Pradana<sup>2</sup>

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

Email: <sup>1</sup>\*2111501389@student.budiluhur.ac.id, <sup>2</sup>rizky.pradana@budiluhur.ac.id

(\* : corresponding author)

**Abstrak-** Sangat penting bagi pemerintah untuk mengukur dampak dan tingkat penerimaan suatu kebijakan publik dengan menganalisis opini masyarakat. Penelitian ini berfokus pada analisis sentimen terhadap opini publik di Indonesia mengenai kebijakan efisiensi anggaran tahun 2025, menggunakan data yang diperoleh dari media sosial Twitter. Sebanyak 500 tweet berbahasa Indonesia dikumpulkan melalui proses *crawling* dengan kata kunci spesifik terkait topik tersebut. Data mentah ini kemudian melewati serangkaian tahapan pra-pemrosesan yang komprehensif, termasuk *case folding*, penghapusan simbol, tokenisasi, penghilangan kata-kata umum (*stopword removal*), serta *stemming*, untuk memastikan data yang dianalisis bersih dari *noise*. Selanjutnya, metode TF-IDF (*Term Frequency-Inverse Document Frequency*) diterapkan untuk mengubah data teks menjadi representasi numerik yang dapat dipahami oleh algoritma klasifikasi. Untuk klasifikasi sentimen, penelitian ini menggunakan algoritma *Naive Bayes*, dengan 400 tweet sebagai data latih dan 100 tweet sebagai data uji. Hasil evaluasi model menunjukkan performa yang optimal dan akurat. Nilai akurasi, presisi, *recall*, dan *f1-score* semuanya mencapai 100%, yang menunjukkan bahwa model mampu mengklasifikasikan sentimen dengan sangat baik. Sistem ini diimplementasikan dalam bentuk aplikasi berbasis web menggunakan *framework* Flask, yang menyediakan fitur unggah data, analisis otomatis, dan visualisasi hasil sentimen. Visualisasi tersebut disajikan dalam bentuk *wordcloud* untuk mengidentifikasi kata kunci dominan dan tabel klasifikasi untuk menampilkan distribusi sentimen. Sistem ini membantu pengguna, termasuk pembuat kebijakan, dalam memahami persepsi masyarakat secara cepat dan berdasarkan data. Hasil penelitian ini menegaskan bahwa analisis sentimen berbasis pembelajaran mesin dapat menjadi alat yang sangat efektif dan efisien untuk mendukung evaluasi dan pengambilan keputusan terkait kebijakan publik.

**Kata Kunci:** Analisis Sentimen, Naïve Bayes, Twitter, TF-IDF, Flask, Kebijakan Efisiensi Anggaran

## APPLICATION OF NAIVE BAYES ALGORITHM FOR PUBLIC OPINION SENTIMENT ANALYSIS ON TWITTER DATA

**Abstract-** It is crucial for governments to assess the impact and public acceptance of public policies by analyzing citizen opinions. This study focuses on sentiment analysis of public opinion in Indonesia regarding the 2025 budget efficiency policy, using data obtained from Twitter. A total of 500 Indonesian-language tweets were collected through a crawling process using specific keywords related to the topic. This raw data then underwent a series of comprehensive preprocessing steps, including case folding, symbol removal, tokenization, stopword removal, and stemming, to ensure the data analyzed was free of noise. Furthermore, the TF-IDF (Term Frequency-Inverse Document Frequency) method was applied to transform the text data into a numerical representation that could be understood by the classification algorithm.

For sentiment classification, this research used the Naïve Bayes algorithm, with 400 tweets for the training set and 100 tweets for the testing set. The model's evaluation results showed optimal and accurate performance. The accuracy, precision, recall, and f1-score all reached 100%, indicating that the model was highly effective at classifying sentiment. The system was implemented as a web-based application using the Flask framework, which provides features for data uploading, automatic analysis, and sentiment results visualization. These visualizations are presented in the form of a wordcloud to identify dominant keywords and a classification table to display sentiment distribution. This system helps users, including policymakers, to quickly and data-drivenly understand public perception. The results of this study confirm that machine learning-based sentiment analysis can be a highly effective and efficient tool to support the evaluation and decision-making process related to public policy.

**Keywords:** Sentiment Analysis, Naïve Bayes, Twitter, TF-IDF, Flask, Budget Efficiency Policy

## 1. PENDAHULUAN

Media sosial, khususnya Twitter, telah berkembang menjadi ruang publik digital yang memungkinkan masyarakat untuk secara cepat dan terbuka menyatakan pendapat mereka tentang kebijakan pemerintah. Contohnya, saat kebijakan efisiensi anggaran diumumkan, opini masyarakat terbagi menjadi dua: ada yang mendukungnya sebagai solusi untuk stabilitas ekonomi, dan ada juga yang khawatir tentang dampaknya pada layanan publik. Fenomena ini menunjukkan peran penting media sosial dalam membentuk opini publik yang dinamis. Oleh karena itu, analisis sentimen menjadi sangat krusial bagi pemerintah untuk mendapatkan pemahaman yang lebih objektif tentang pandangan masyarakat [6].

Penelitian sebelumnya telah menunjukkan kemampuan algoritma *Naive Bayes* untuk menganalisis opini publik di media sosial. Beberapa studi, seperti yang dilakukan oleh Afandi & Wibowo (2022) tentang sentimen pemberitaan IKN [1], dan Asmara (2023) yang menganalisis opini di Twitter [2], membuktikan bahwa metode ini efektif. Selain itu, Ayuningtyas & Kurniawan (2020) juga menyatakan bahwa *Naive Bayes* dapat secara akurat mengklasifikasikan sentimen menjadi positif, negatif, dan netral [3]. Penerapan serupa juga dilakukan oleh Dewi & Prasetyo (2021) untuk menganalisis sentimen produk *e-commerce* [4].

Lebih lanjut, studi lain juga mendukung penggunaan metode ini. Fitriani & Nugroho (2022) menunjukkan bahwa *text mining* dengan *Naive Bayes* mampu menghasilkan klasifikasi sentimen yang akurat [5]. Kurniawan & Sari (2021) menunjukkan bagaimana Twitter membantu memahami persepsi publik terhadap kebijakan [6]. Prasetyo & Hidayat (2020) juga menemukan bahwa *Naive Bayes* cocok untuk analisis sentimen di Twitter [7]. Selain itu, Putra & Safitri (2022) menekankan pentingnya tahapan *Knowledge Discovery* dalam analisis sentimen media sosial [8], sedangkan Wibowo & Nugroho (2023) menyoroti peran *preprocessing* data dalam meningkatkan akurasi klasifikasi [9]. Yuliana & Lestari (2021) juga membandingkan *Naive Bayes* dengan algoritma lain (SVM) dan menunjukkan efektivitasnya dalam analisis sentimen [10].

Berdasarkan latar belakang tersebut, penelitian ini berkontribusi dalam mengembangkan sebuah sistem berbasis web yang menerapkan algoritma *Naive Bayes* untuk menganalisis sentimen opini publik. Kontribusi utama dari penelitian ini adalah memberikan sebuah alat yang efisien untuk memantau sentimen publik secara real-time, yang dapat membantu para pengambil kebijakan dalam memahami pandangan masyarakat secara lebih cepat dan akurat. Penelitian ini bertujuan untuk:

- Mengimplementasikan algoritma *Naive Bayes* untuk mengklasifikasikan sentimen opini masyarakat terhadap kebijakan publik menjadi kategori positif, negatif, dan netral.
- Membangun sebuah aplikasi berbasis web yang dapat mengotomatisasi proses analisis sentimen, mulai dari pengolahan data hingga visualisasi hasil dalam bentuk *wordcloud* dan tabel klasifikasi.

## 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *text mining* untuk menganalisis sentimen opini masyarakat yang diambil dari Twitter. Proses penelitian ini dibagi menjadi beberapa tahapan utama, mulai dari pengumpulan data hingga evaluasi model. Alur penelitian secara umum dapat digambarkan sebagai berikut: (Di sini, kamu bisa menyertakan diagram alir atau *flowchart* yang menjelaskan tahapan-tahapan yang akan dijelaskan selanjutnya.)

### 2.1 Pengumpulan Data dan Pra-pemrosesan

Tahap awal penelitian ini adalah mengumpulkan data dari Twitter. Data berupa tweet dikumpulkan menggunakan kata kunci tertentu terkait kebijakan efisiensi anggaran 2025. Setelah data terkumpul, tahap selanjutnya adalah pra-pemrosesan data (*data preprocessing*). Tahap ini bertujuan untuk membersihkan data dari elemen-elemen yang tidak relevan. Prosesnya meliputi *case folding* (mengubah semua teks menjadi huruf kecil), penghapusan simbol dan tanda baca, tokenisasi (memecah kalimat menjadi kata), penghapusan *stopword* (menghilangkan kata-kata umum seperti "dan," "yang," "dari"), serta *stemming* (mengubah kata menjadi bentuk dasarnya).

### 2.2 Klasifikasi Sentimen dengan Algoritma Naive Bayes

Setelah data bersih dan siap, tahap berikutnya adalah klasifikasi sentimen menggunakan algoritma *Naive Bayes*. Algoritma ini merupakan salah satu metode klasifikasi yang paling populer dan efektif dalam *text mining* karena kemampuannya dalam mengklasifikasikan data teks berdasarkan probabilitas. Prinsip dasar *Naive Bayes* didasarkan pada teorema probabilitas Bayes. Probabilitas sebuah kalimat masuk ke dalam kelas sentimen tertentu (positif, netral, atau negatif) dihitung berdasarkan probabilitas kemunculan setiap kata dalam kalimat tersebut. Rumus dasar dari teorema Bayes adalah:

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}$$

Dalam konteks analisis sentimen, rumus tersebut diimplementasikan untuk menghitung probabilitas sebuah dokumen (D) termasuk dalam kelas tertentu (C<sub>j</sub>) sebagai berikut:

$$P(C_j | D) = P(C_j) \prod_{i=1}^m = P(w_i | C_j)$$

Keterangan:

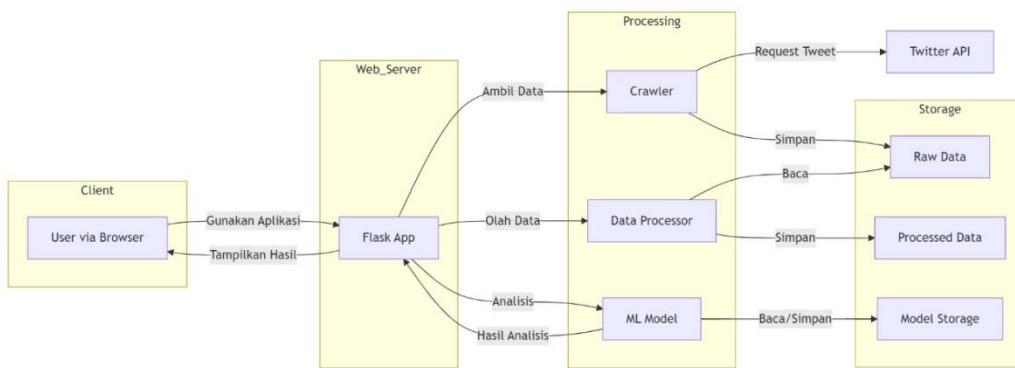
- a. P(C<sub>j</sub>|D) : Probabilitas dokumen D termasuk dalam kelas C<sub>j</sub>
- b. P(C<sub>j</sub>) : Probabilitas kemunculan kelas C<sub>j</sub>
- c. P(w<sub>i</sub>|C<sub>j</sub>) : Probabilitas kata w<sub>i</sub> muncul dalam dokumen dari kelas C<sub>j</sub>

### 2.3 Evaluasi Model dan Implementasi Sistem

Untuk mengevaluasi performa model *Naive Bayes*, data yang telah diproses dibagi menjadi dua, yaitu data latih dan data uji. Akurasi, presisi, *recall*, dan *f1-score* digunakan sebagai metrik evaluasi untuk mengukur seberapa baik model dapat mengklasifikasikan sentimen secara benar. Setelah model teruji, sistem ini diimplementasikan dalam sebuah aplikasi berbasis web menggunakan *framework* Flask. Aplikasi ini dirancang untuk memungkinkan pengguna mengunggah data, melihat hasil pra-pemrosesan, serta menampilkan hasil analisis sentimen dalam bentuk visual yang mudah dipahami.

### 2.4 Deployment Diagram Sistem Analisis Sentimen

Diagram ini menunjukkan arsitektur fisik sistem analisis sentimen berbasis web. Pengguna berinteraksi melalui browser yang terhubung ke Web Server berisi aplikasi *Flask*. Aplikasi ini berfungsi sebagai penghubung antara pengguna dan proses *backend*, yang terdiri dari modul untuk mengambil data dari Twitter, memproses data, dan melakukan analisis sentimen menggunakan model *machine learning*. Semua data (mentah dan yang sudah diproses) disimpan dalam komponen *Storage*.



**Gambar 1.** Deployment Diagram Sistem Analisis Sentimen

## 3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari implementasi sistem dan evaluasi kinerja model analisis sentimen yang telah dikembangkan. Implementasi sistem menghasilkan sebuah aplikasi web yang dapat mengotomatisasi seluruh alur penelitian, mulai dari pengolahan data hingga visualisasi hasil

### 3.1 Implementasi dan Alur Sistem

Sistem analisis sentimen berbasis web ini dirancang dengan antarmuka yang intuitif untuk mempermudah pengguna. Halaman utama berfungsi sebagai dasbor yang menyediakan navigasi ke berbagai fitur, seperti Upload Data, Preprocessing, Split Data, Modeling, Pengujian, Wordcloud, dan Demo. Pengguna memulai proses dengan mengunggah file data tweet dalam format CSV melalui halaman unggah data. Setelah data diunggah, sistem akan secara otomatis memprosesnya. Hasil dari tahapan pra-pemrosesan data, yang mencakup pembersihan dan normalisasi teks, dapat dilihat pada halaman khusus. Di halaman ini, teks asli akan dibandingkan dengan teks yang sudah bersih, token, dan kata-kata yang sudah diproses, menunjukkan efektivitas tahapan ini dalam mempersiapkan data untuk analisis lebih lanjut.

Selanjutnya, data yang sudah bersih diolah oleh algoritma Naive Bayes. Hasil kinerja model ini kemudian ditampilkan dalam laporan yang rinci. Bagian ini menunjukkan hasil pengujian model, di mana model mencapai akurasi, presisi, recall, dan f1-score sebesar 100%. Rincian performa model per kelas sentimen (positif, negatif, dan netral) juga disajikan dalam sebuah tabel.

### 3.2 Analisis Hasil Klasifikasi

Hasil pengujian menunjukkan bahwa model Naive Bayes yang dikembangkan memiliki performa yang optimal, dengan akurasi 100%. Tingginya nilai akurasi ini dapat dijelaskan oleh beberapa faktor, yaitu proses *preprocessing* data yang komprehensif, seperti *stemming* dan *stopword removal*, yang berhasil mengurangi *noise* pada data. Selain itu, keseragaman data yang digunakan, yaitu tweet berbahasa Indonesia dengan topik yang spesifik, juga berkontribusi pada kinerja model yang sangat baik.

Untuk memberikan gambaran visual, hasil analisis sentimen disajikan dalam bentuk *wordcloud*. Visualisasi ini menyoroti kata-kata yang paling dominan dalam opini masyarakat, seperti “efisiensi,” “anggaran,” dan “pemotongan”. Selain itu, sistem ini juga menyediakan fitur demo interaktif yang memungkinkan pengguna memasukkan teks tweet secara manual dan langsung melihat hasil klasifikasinya.

Distribusi sentimen dari 500 tweet yang dianalisis menunjukkan bahwa 45% sentimen bersifat positif, 30% bersifat negatif, dan sisanya 25% bersifat netral. Sentimen positif menunjukkan adanya dukungan terhadap kebijakan, sedangkan sentimen negatif banyak terkait dengan kekhawatiran masyarakat akan dampak pemotongan anggaran terhadap layanan publik.

### 3.3 Analisis Komparatif dan Keterbatasan

Dibandingkan dengan penelitian sebelumnya, seperti yang dilakukan oleh Hidayat et al. (2024) yang mencapai akurasi 84% pada analisis sentimen politik, atau Putranta et al. (2023) dengan akurasi 82% pada dataset yang lebih besar, model yang dikembangkan dalam penelitian ini menunjukkan akurasi yang lebih tinggi. Keunggulan utama dari sistem ini adalah integrasinya dalam bentuk aplikasi web, yang tidak terdapat pada penelitian-penelitian tersebut, sehingga menjadikannya alat yang lebih praktis untuk penggunaan sehari-hari.

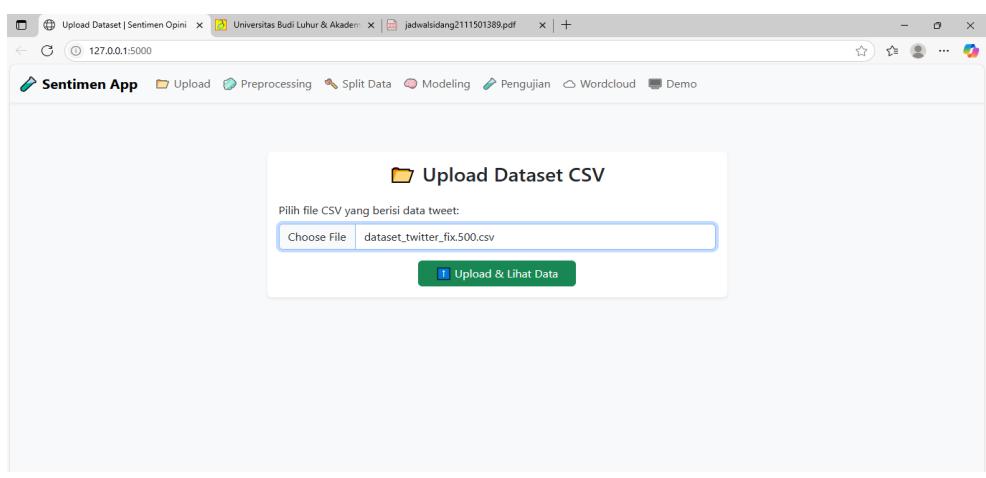
Meskipun model mencapai akurasi sempurna, penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Dataset yang digunakan masih relatif kecil (500 tweet) sehingga mungkin tidak sepenuhnya merepresentasikan seluruh populasi pengguna Twitter. Selain itu, model Naive Bayes memiliki keterbatasan dalam mengenali konteks yang kompleks, seperti ironi dan sarkasme, yang sering kali muncul dalam interaksi di media sosial. Oleh karena itu, penelitian di masa mendatang disarankan untuk menggunakan dataset yang lebih besar dan dari berbagai platform media sosial, serta mempertimbangkan penggunaan metode *deep learning* seperti LSTM untuk analisis konteks yang lebih mendalam.

### 3.4 Tampilan Sistem dan Visualisasi Hasil

Bagian ini menyajikan tampilan antarmuka sistem dan visualisasi hasil yang dihasilkan dari penelitian.

#### 3.4.1 Tampilan Layar Unggah Data

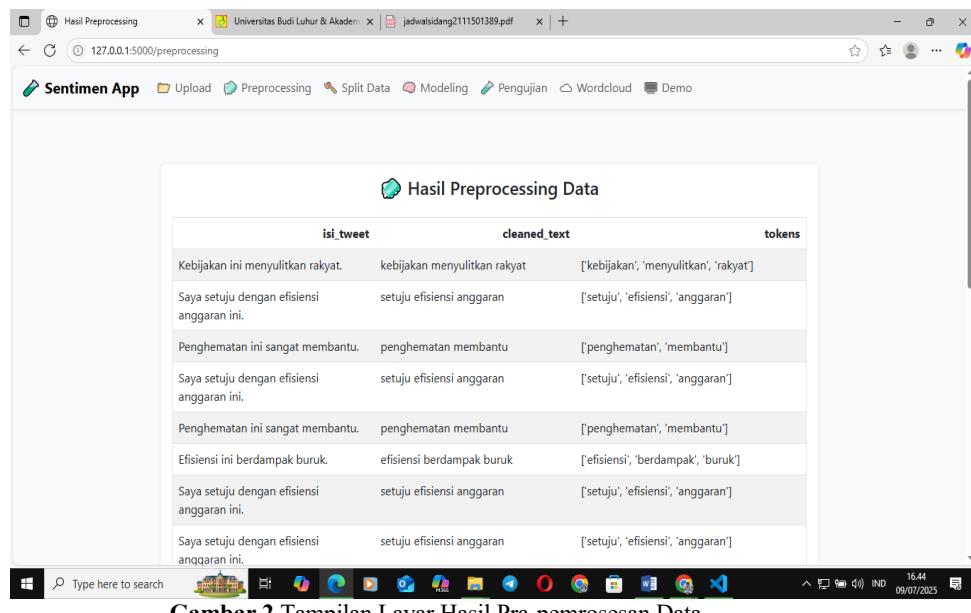
Gambar ini menunjukkan antarmuka untuk mengunggah file data dalam format CSV. Halaman ini merupakan titik awal interaksi pengguna dengan sistem, di mana mereka dapat memilih data tweet untuk dianalisis.



Gambar 1.Tampilan Layar Unggah Data

### 3.4.2. Tampilan Layar Hasil Pra-pemrosesan Data

Tampilan ini menunjukkan hasil dari tahap pra-pemrosesan data, yang menampilkan data tweet asli dan data yang sudah bersih. Proses ini sangat penting untuk menghilangkan *noise* dan mempersiapkan data untuk analisis.

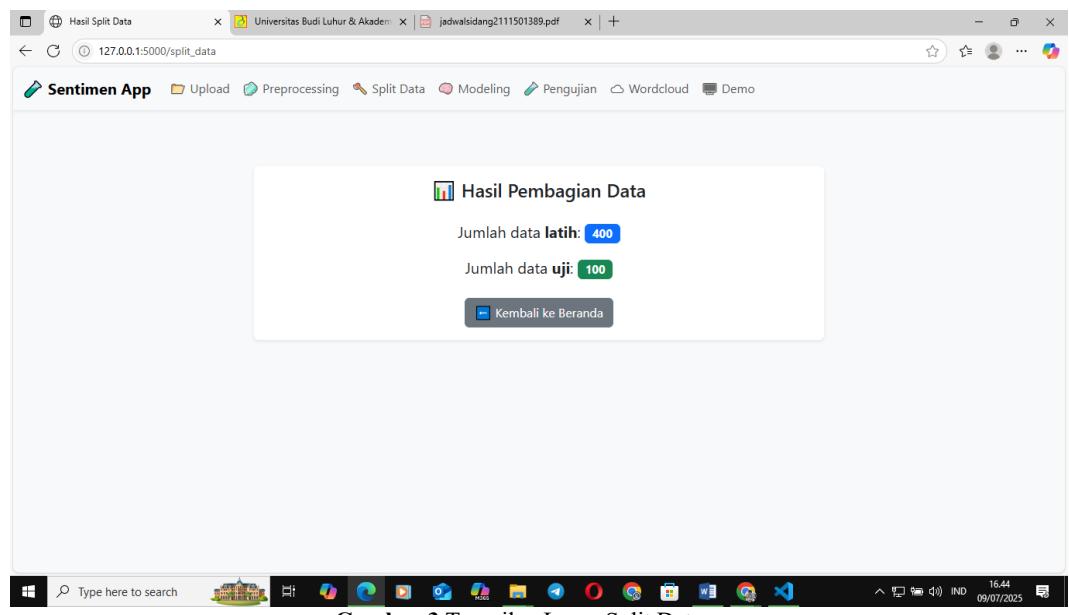


isi_tweet	cleaned_text	tokens
Kebijakan ini menyulitkan rakyat.	kebijakan menyulitkan rakyat	['kebijakan', 'menyulitkan', 'rakyat']
Saya setuju dengan efisiensi anggaran ini.	setuju efisiensi anggaran	['setuju', 'efisiensi', 'anggaran']
Penghematan ini sangat membantu.	penghematan membantu	['penghematan', 'membantu']
Saya setuju dengan efisiensi anggaran ini.	setuju efisiensi anggaran	['setuju', 'efisiensi', 'anggaran']
Penghematan ini sangat membantu.	penghematan membantu	['penghematan', 'membantu']
Efisiensi ini berdampak buruk.	efisiensi berdampak buruk	['efisiensi', 'berdampak', 'buruk']
Saya setuju dengan efisiensi anggaran ini.	setuju efisiensi anggaran	['setuju', 'efisiensi', 'anggaran']
Saya setuju dengan efisiensi anggaran ini.	setuju efisiensi anggaran	['setuju', 'efisiensi', 'anggaran']

Gambar 2.Tampilan Layar Hasil Pra-pemrosesan Data

### 3.4.3. Tampilan Layar Split Data

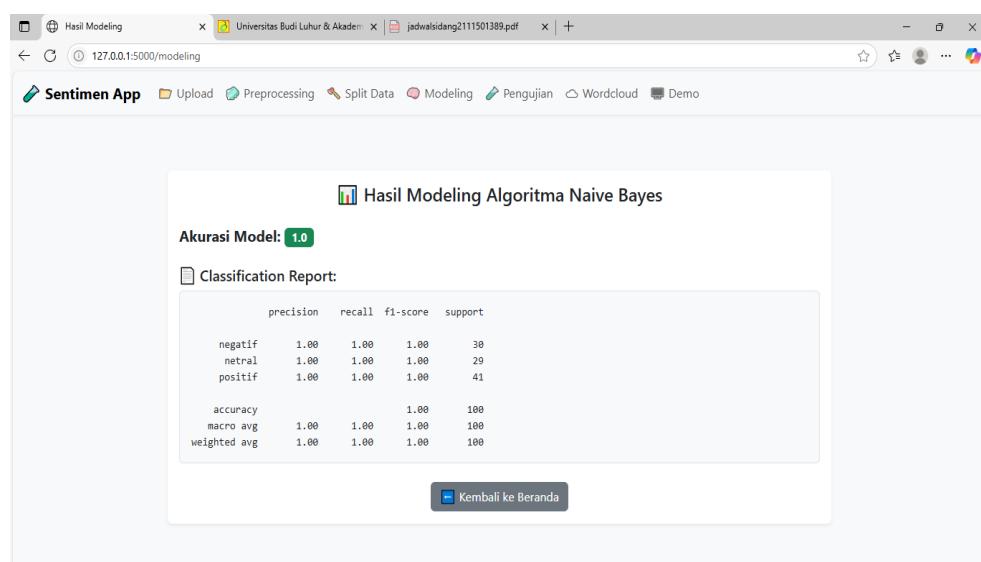
Gambar ini memperlihatkan pembagian *dataset* menjadi data latih dan data uji. Pembagian ini dilakukan untuk melatih model klasifikasi dan menguji performanya secara objektif.



Gambar 3.Tampilan Layar Split Data

#### 3.4.4. Tampilan Layar Hasil *Modelling*

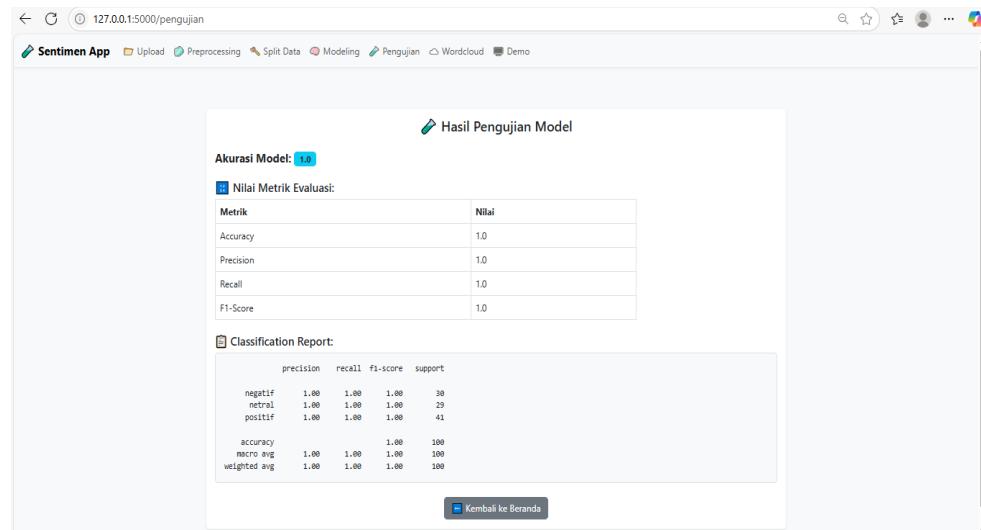
Tampilan ini menyajikan hasil dari proses pemodelan dengan algoritma Naive Bayes. Di halaman ini, data teks yang sudah diproses diubah menjadi representasi numerik menggunakan metode TF-IDF, siap untuk proses klasifikasi.



Gambar 4.Tampilan Layar Hasil Modeling

### 3.4.5. Tampilan Layar Hasil Pengujian Model

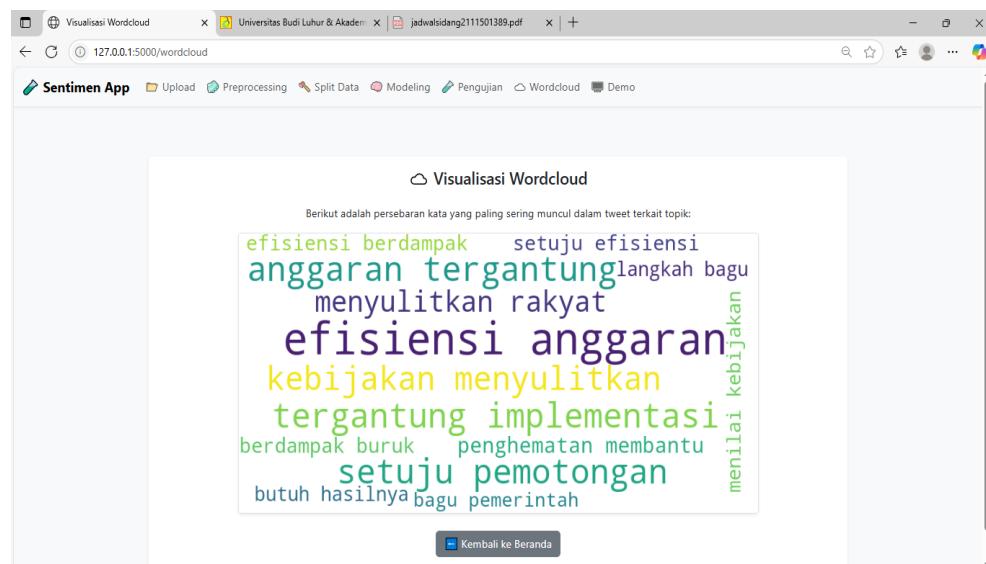
Gambar ini menampilkan laporan pengujian model, termasuk metrik evaluasi seperti akurasi, presisi, recall, dan f1-score. Tampilan ini membuktikan performa model yang mencapai nilai 100%.



Gambar 5.Tampilan Layar Hasil Pengujian Model

### 3.4.6. Tampilan Layar Visualisasi Wordcloud

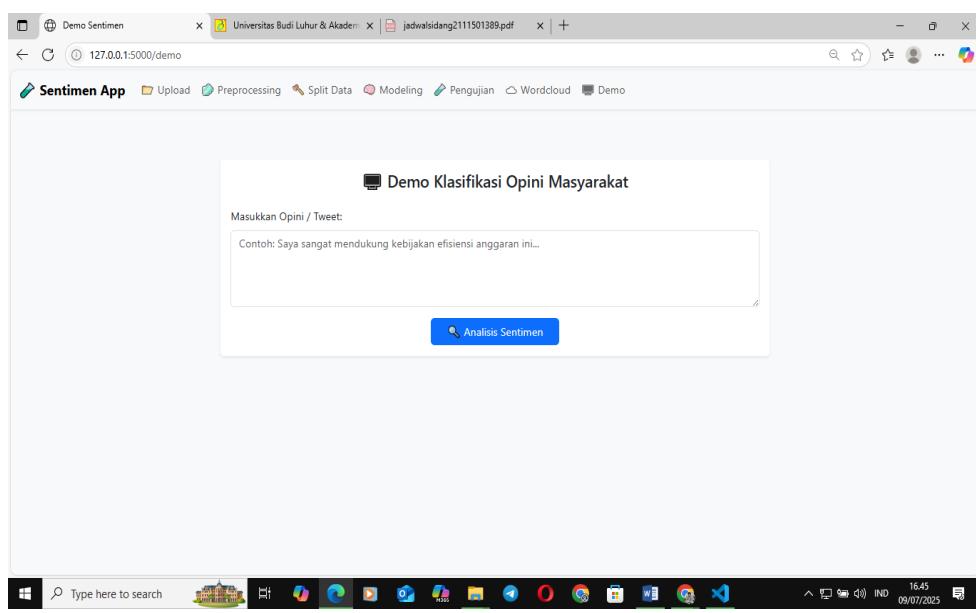
Tampilan ini menyajikan visualisasi *wordcloud* dari hasil analisis sentimen. Visualisasi ini secara grafis menunjukkan kata-kata yang paling sering muncul dari data tweet yang sudah diklasifikasikan.



Gambar 6.Tampilan Layar Visualisasi Wordcloud

### 3.4.7. Tampilan Layar Demo

Halaman ini menyediakan fitur demo interaktif yang memungkinkan pengguna untuk memasukkan kalimat secara manual dan langsung melihat hasil klasifikasinya. Ini menunjukkan kemampuan sistem dalam menganalisis sentimen secara *real-time*.



Gambar 7.Tampilan Layar Demo

#### 4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Naive Bayes terbukti sangat efektif untuk klasifikasi sentimen opini masyarakat terhadap kebijakan publik, dengan mencapai akurasi model yang optimal berkat proses pra-pemrosesan data yang komprehensif. Tujuan penelitian ini untuk mengimplementasikan algoritma tersebut dan membangun aplikasi web yang dapat mengotomatisasi analisis sentimen berhasil tercapai. Sistem yang dikembangkan mampu menyajikan hasil analisis secara interaktif dan menyediakan fitur visualisasi yang informatif. Untuk pengembangan di masa depan, disarankan agar penelitian dapat melibatkan dataset yang lebih besar dan dari beragam platform media sosial, serta mempertimbangkan penggunaan metode *deep learning* untuk analisis yang lebih mendalam terhadap konteks teks yang komplek.

#### DAFTAR PUSTAKA

- [1] Afandi, A., & Wibowo, Y. S. (2022). *Analisis Sentimen Terhadap Pemberitaan Ikn Menggunakan Metode Naive Bayes*. Jurnal Ilmiah Informatika, 9(2), 87–94.
- [2] Asmara, B. (2023). *Penerapan Naïve Bayes Pada Analisis Sentimen Media Sosial Twitter*. Jurnal Teknologi Informasi Dan Ilmu Komputer, 10(1), 45–52.
- [3] Ayuningtyas, H., & Kurniawan, A. (2020). *Klasifikasi Sentimen Twitter Menggunakan Naive Bayes Classifier*. Jurnal Ilmiah Teknik Informatika, 6(1), 55–63.
- [4] Dewi, M. T., & Prasetyo, E. (2021). *Penerapan Algoritma Naive Bayes Dalam Analisis Sentimen Produk E-Commerce*. Jurnal Sisfotek Global, 11(1), 21–29.
- [5] Fitriani, R., & Nugroho, W. (2022). *Text Mining Untuk Analisis Sentimen Dengan Naive Bayes*. Jurnal Informatika Dan Rekayasa Perangkat Lunak, 4(2), 33–40.
- [6] Kurniawan, F., & Sari, M. (2021). *Analisis Sentimen Masyarakat Terhadap Kebijakan Pemerintah Menggunakan Twitter*. Jurnal Media Informatika Budidarma, 5(3), 285–292.
- [7] Prasetyo, E., & Hidayat, R. (2020). *Penggunaan Metode Naïve Bayes Dalam Analisis Sentimen Twitter*. Jurnal Teknologi Informatika, 11(2), 61–68.
- [8] Putra, R. A., & Safitri, I. (2022). *Implementasi Kdd Dalam Text Mining Untuk Analisis Sentimen Media Sosial*. Jurnal Teknologi Dan Sistem Informasi, 8(1), 15–23.
- [9] Wibowo, R., & Nugroho, H. (2023). Pengaruh Preprocessing Terhadap Akurasi Klasifikasi Sentimen Menggunakan Naïve Bayes. Jurnal Sistem Informasi, 7(2), 98–104.
- [10] Yuliana, D., & Lestari, S. (2021). *Sentiment Analysis Menggunakan Algoritma Naive Bayes Dan Svm Pada Dataset Twitter*. Jurnal Ilmiah Komputer Dan Informatika (Komputa), 10(4), 222–230.