

# Analisis Sentimen Kebijakan Insentif Mobil Listrik Menggunakan TF-IDF dan Naïve Bayes

Yande Pramana Yustika Pradeva<sup>a1</sup>, I Made Widiartha<sup>a2</sup>, I Putu Satwika<sup>a3</sup>

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>pradeva.2308561125@student.unud.ac.id  
<sup>2</sup>madewidiartha@email.com  
<sup>3</sup>satwika@unud.ac.id

## Abstract

*This study aims to analyze public sentiment regarding the Indonesian government's electric vehicle (EV) incentive policy using YouTube comments as the data source. The research applies text preprocessing steps including cleaning, normalization, stopword removal, tokenization, and stemming to prepare the textual data. The cleaned data is transformed into numerical representation using the Term Frequency-Inverse Document Frequency (TF-IDF) method and classified using the Multinomial Naïve Bayes algorithm. To address class imbalance in the dataset, Synthetic Minority Over-sampling Technique (SMOTE) is applied. The model evaluation metrics include accuracy, precision, recall, and F1-score. Based on the evaluation, the model achieves an accuracy of 71%. The model performs better in classifying negative comments, as shown by a higher recall and F1-score in the negative class compared to the positive class. These findings indicate that public responses to the EV incentive policy tend to be more critical. This study provides insights into public opinion that can serve as a valuable reference for policymakers in designing more effective and well-communicated incentive strategies for promoting electric vehicle adoption in Indonesia.*

**Keywords:** Insentif Mobil Listrik, Naïve Bayes, TF-IDF, SMOTE, Filter Kata, Analisis Sentimen, Dataset Kaggle.

## 1. Pendahuluan

Di tengah kemajuan teknologi modern, berbagai sektor kehidupan mengalami transformasi signifikan, termasuk sektor otomotif. Salah satu isu yang kini menarik perhatian publik adalah kebijakan insentif pemerintah terhadap mobil listrik. Insentif ini diberikan sebagai bagian dari upaya percepatan transisi menuju kendaraan ramah lingkungan. Dengan adanya mobil listrik diharapkan dapat mengurangi pencemaran bahan bakar fosil terhadap lingkungan. General Manager PLN UNIT Induk Distribusi Sulawesi Selatan Moch Andy Adchaminoerdin pada [kompas.com](https://www.kompas.com), 2024 menyampaikan bahwa penggunaan kendaraan listrik berkontribusi mengurangi 56 persen emisi karbon. Meskipun mobil listrik menawarkan efisiensi energi dan manfaat bagi lingkungan, respons masyarakat terhadap kebijakan insentif tersebut masih beragam.

Penelitian ini difokuskan pada analisis sentimen masyarakat terhadap insentif mobil listrik, bukan terhadap teknologi mobil listrik secara umum. Pemahaman terhadap opini publik mengenai insentif tersebut penting untuk menilai efektivitas kebijakan serta mengetahui persepsi masyarakat apakah mereka mendukung atau justru skeptis terhadap program tersebut. Salah satu tantangan utama dalam analisis ini adalah minimnya pemahaman publik mengenai bentuk dan manfaat dari insentif yang diberikan, serta masih terbatasnya penyebaran informasi yang akurat. Hal ini dapat memengaruhi bagaimana masyarakat menanggapi program tersebut, baik secara positif maupun negatif. Selain itu, isu-isu seperti keterbatasan infrastruktur pengisian daya, harga mobil listrik yang tinggi, dan persepsi terhadap ketimpangan akses juga turut mempengaruhi sentimen publik.

Untuk memperoleh gambaran yang lebih mendalam mengenai persepsi publik, penelitian ini menggunakan pendekatan analisis sentimen dengan memanfaatkan komentar-komentar dari pengguna YouTube. Platform ini dipilih karena memiliki tingkat partisipasi pengguna yang tinggi serta menyajikan ruang diskusi terbuka pada berbagai isu aktual, termasuk kebijakan pemerintah terkait insentif kendaraan listrik. Video-video yang membahas tema tersebut sering kali menjadi wadah bagi masyarakat untuk menyuarakan opini, baik berupa dukungan, kritik, maupun pandangan netral. Oleh karena itu, komentar YouTube dinilai sebagai sumber data yang representatif dalam menangkap opini masyarakat secara langsung dan organik terhadap kebijakan insentif mobil listrik.

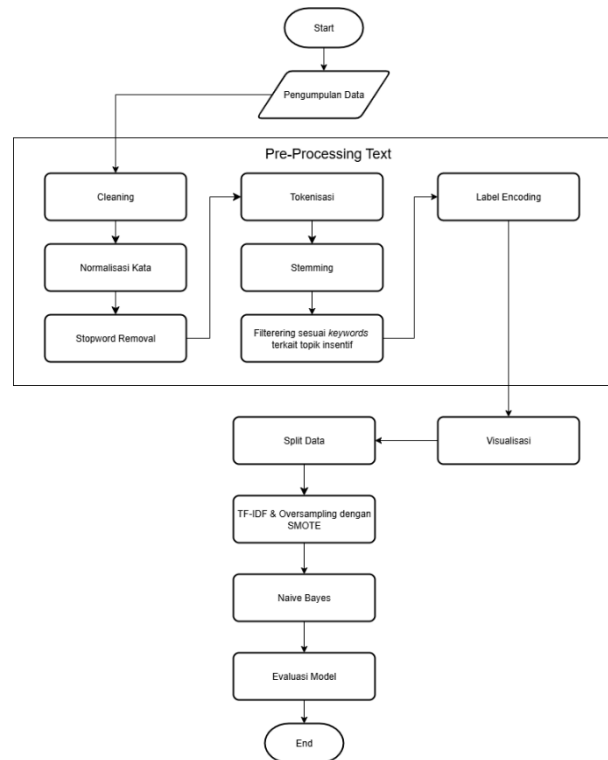
Dalam proses klasifikasi sentimen, digunakan algoritma Multinomial Naïve Bayes, salah satu metode populer dalam pembelajaran mesin yang mengandalkan pendekatan probabilistik. Naïve Bayes dikenal karena kesederhanaannya namun efektif dalam mengklasifikasikan teks ke dalam kategori positif atau negatif. Algoritma ini mengestimasi probabilitas suatu kategori berdasarkan data sebelumnya dan sangat sesuai untuk diterapkan dalam konteks klasifikasi komentar di platform seperti YouTube.

Melalui penelitian ini, dilakukan klasifikasi komentar masyarakat terkait insentif mobil listrik untuk mengetahui arah opini publik. Dengan menggunakan teknik ekstraksi fitur TF-IDF, serta evaluasi menggunakan metrik seperti akurasi, presisi, recall, dan F1-score, diharapkan dapat diperoleh pemahaman yang lebih mendalam mengenai respon masyarakat terhadap kebijakan insentif kendaraan listrik. Hasil analisis ini diharapkan dapat menjadi acuan bagi pemerintah atau pihak terkait dalam merumuskan kebijakan lanjutan yang lebih tepat sasaran.

## 2. Metode Penelitian

Pada bagian metode penelitian terdapat alur tahapan penelitian yang dirancang dan dijalankan oleh penulis. Dalam studi ini menggunakan data berupa komentar dari media *online* YouTube yang diambil dari beberapa video yang membahas topik terkait Insentif Mobil Listrik di Indonesia. Data ini merupakan data sekunder yang penulis dapatkan dari situs kaggle yang memiliki 1517 baris data dan lima kolom termasuk label, yaitu positif, negatif serta netral. Kemudian data tersebut di *filter* dengan *keywords* terkait topik tersebut yang nanti dianalisis oleh penulis untuk mengetahui opini masyarakat terkait penerapan Insentif Mobil Listrik di Indonesia. Dalam penelitian ini menggunakan algoritma *Naive Bayes* dengan ekstraksi fitur TF-IDF dengan beberapa tahapan seperti yang terdapat pada Gambar 1.

Gambar 1 menggambarkan tahapan penelitian yang meliputi pengumpulan data *sekunder* yang didapat dari situs terbuka kaggle, tahap pemrosesan dan penyeimbangan data dengan *pre-processing* meliputi cleaning, normalisasi, stopword, tokenisasi, stemming, filtering dan ekstraksi label sentimen. Kemudian tahap visualisasi, split data dan pembobotan fitur dengan TF-IDF serta *oversampling* dengan SMOTE untuk terdapat menyeimbangkan data sebelum masuk tahap pemodelan. Setelah itu tahap *modeling* menggunakan Multinomial Naive Bayes. Dan tahap terakhir merupakan tahap evaluasi model yang menghasilkan *output* berupa matriks evaluasi termasuk akurasi, presisi, *recall*, dan F-1 score dan selesai.



**Gambar 1.** Tahapan Metode Penelitian

## 2.1. Pengumpulan Data

*Dataset* pada penelitian ini didapat dari situs terbuka kaggle dengan kata kunci “mobil listrik”. Data yang dikumpulkan dalam format file (.csv) dengan 1517 baris data dengan lima kolom sentimen, dalam data ini terdiri dari tiga jenis sentimen yaitu sentimen positif, negatif dan netral. Data ini adalah *opini* masyarakat yang diambil dari komentar dari berbagai video youtube yang membahas tentang insentif mobil listrik di Indonesia. Sehingga data yang diperoleh adalah data sekunder.

## 2.2. Text Pre-Processing

*Text Pre-Processing* merupakan langkah awal pada pengklasifikasiian teks dengan merubah bentuk teks agar menjadi sesuai sebelum di nantinya diproses pada tahapan berikutnya [1]. Pada tahapan ini meliputi pembersihan, transformasi, dan penataan teks agar nanti dapat diproses dengan baik pada algoritma *Natural Language Processing* (NLP) atau model *machine learning* [1]. Tujuan dari proses ini untuk meningkatkan kualitas data dengan menghapus informasi yang tidak relevan, menyamakan variasi representasi kata, dan menyamakan susunan dan format teks. Adapun teknik yang biasa digunakan dalam tahap ini, yaitu *cleaning*, *lowercase*, normalisasi, tokenisasi, *stopwords*, dan *stemming*. Dalam penelitian ini, tahap *pre-processing* hanya melibatkan *cleaning*, normalisasi, tokenisasi, *stopwords*, dan *stemming* menyesuaikan karakteristik dari data.

### a. Cleaning

*Cleaning* adalah tahap awal pada pemrosesan teks yang terfokus pada pembersihan dari berbagai elemen-elemen maupun simbol yang tidak penting [1]. Proses ini terdiri dari penghapusan data yang duplikat, angka, simbol, *emoji*, serta tanda baca. Pada penelitian ini dilakukan penghapusan data kosong dan atribut yang tidak diperlukan seperti, id komentar, nama akun dan tanggal.

**b. Normalisasi**

Normalisasi adalah tahapan yang mengubah kata-kata tidak baku, istilah gaul atau kurang sesuai menjadi kata yang baku sesuai kaidah bahasa Indonesia yang tercantum dalam Kamus Besar Bahasa Indonesia (KBBI) [1]. Proses ini terdiri dari penggantian kata singkatan, kata tidak baku, istilah gaul untuk menjadikannya konsisten yaitu baku. Normalisasi ini agar mempermudah proses selanjutnya. Pada penelitian ini dilakukan Normalisasi dengan menggunakan kamus kata baku dan tidak baku yang diperoleh dari media kaggle dengan kata kunci "kamus slag", yang berisi kumpulan kata-kata tidak baku dan baku baik yang tersingkat maupun tidak dengan format file (.xlsx) dan dengan penambahan penyesuaian kata-kata baru menyesuaikan data yang kita proses yang tidak dimuat di kamus kata.

**c. Stopword Remover**

*Stopword* adalah tahapan untuk menghapus kata-kata yang tidak relevan di dalam teks [1]. Dalam konteks analisis seperti kata sambung, kata waktu, atau kata fungsional lainnya dihapus karena kurang relevan. Dengan menghapus kata-kata tersebut hal ini membuat teks menjadi jauh lebih singkat dan fokus pada informasi yang relevan. Pada penelitian ini penghapusan kata-kata tersebut menggunakan *Stopword* dengan Sastrawi yang merupakan *library* dari bahasa pemrograman *python*.

**d. Tokenisasi**

Tokenisasi adalah proses memecah teks menjadi bagian-bagian kecil yang disebut token [1]. Token pada data ini dapat berupa kata maupun frasa menyesuaikan kebutuhan analisis. Pada tahap ini membuat kata menjadi terstruktur dan untuk diproses pada tahap selanjutnya. Pada penelitian ini tokenisasi dilakukan untuk memecah kata-kata menjadi bagian-bagian kecil yang dipisahkan oleh tanda koma.

**e. Stemming**

Stemming adalah tahap merubah kata-kata yang memiliki awalan, imbuhan, akhiran, dan lainnya menjadi bentuk dasarnya [1]. Dengan Tujuan mengubah kata-kata yang memiliki kata dasar yang sama diidentifikasi sebagai sama. Sehingga membantu menyederhanakan representasi data yang meningkatkan efektifitas analisis teks.

**f. Label Encoding**

Label Encoding merupakan teknik merubah nilai dari nilai dari atribut sebuah tabel untuk menjadikan nilai numerik [2][3]. Dalam kasus ini adalah merubah nilai dari atribut sentimen menjadi nilai numerik atau angka, seperti angka nol untuk sentimen negatif, satu untuk sentimen positif dan dua untuk sentimen netral. Hal ini dilakukan untuk membantu dalam proses selanjutnya yaitu pemodelan oleh mesin karena beberapa mesin akan lebih baik menerima fitur dengan data numerik [3].

**2.3. Ekstraksi Fitur *Term Frequency Inverse Document* (TF-IDF)**

Ekstraksi fitur menggunakan TF-IDF merupakan salah satu teknik pembobotan yang kerap diterapkan dalam *information retrieval*, khususnya dalam analisis teks [4]. TF-IDF ini sangatlah penting karena memberikan nilai kontribusi yang tinggi pada kata-kata yang sering muncul di dalam satu dokumen, tapi bobotnya berkurang jika kata itu juga sering muncul di dokumen lain. TF-IDF ini tersusun atas dua komponen utama, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF).

Term Frequency (TF) menggambarkan seberapa sering suatu kata muncul pada sebuah dokumen. Semakin sering muncul maka semakin besar bobotnya. Secara matematis,  $f(t_k, d_j)$  menggambarkan jumlah kemunculan kata  $k$  pada dokumen  $j$ , dan rumus TF dituliskan sebagai berikut[]

$$TF(t_k, d_j) = f(t_k, d_j) \quad (1)$$

Sedangkan, Inverse Document Frequency (IDF) digunakan untuk mengukur sejauh mana suatu

kata terdistribusi dalam kumpulan dokumen. Jika suatu kata muncul hampir di semua dokumen, maka nilainya akan semakin kecil karena dianggap kurang relevan. IDF dihitung dengan membagi jumlah seluruh dokumen  $D$  dengan jumlah dokumen yang mengandung kata tertentu  $df(t)$ , lalu diambil logaritmanya. Rumusnya dituliskan sebagai berikut[]

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

Dan gabungan dari TF dan IDF menghasilkan nilai bobot TF-IDF, yang rumusnya:

$$TF-IDF(t_k, d_j) = TF(t_k, d_j) * IDF(t_k) \quad (3)$$

Metode ini membantu dalam mengatasi pentingnya suatu kata pada setiap dokumen terhadap kategori kelas yang dianalisis.

#### 2.4. Oversampling SMOTE

*Synthetic Minority Over-sampling Technique (SMOTE)* merupakan teknik untuk mengatasi data yang tidak seimbang di dalam proses pembangunan *Machine Learning (ML)* [5]. Oversampling ini digunakan apabila salah satu kelas dalam data memiliki data yang jauh lebih banyak maupun jauh lebih sedikit jika dibandingkan pada kelasnya. Oversampling disini merupakan mengutamakan kelas yang memiliki data lebih sedikit untuk menyamakan data dengan kelas yang lebih banyak. Tujuannya agar model yang dibangun ini tidak bias terhadap salah satu data saja.

#### 2.5. Multinomial Naive Bayes

*Multinomial Naive Bayes* merupakan algoritma *Machine Learning (ML)* yang termasuk kategori supervised learning [4]. Algoritma ini umumnya digunakan dalam analisis data teks dengan menghitung probabilitas sebuah dokumen termasuk ke dalam kelas tertentu. Adapun persamaan dari algoritma ini, yaitu

$$P(C) = \frac{N_c}{N} \quad (4)$$

$N_c$  merupakan kelas  $C$  untuk semua data dan  $N$  merupakan jumlah dari data.

#### 2.6. Evaluasi

Evaluasi model yang dibuat dilakukan agar mengetahui seberapa baik performa model yang dihasilkan. Metode evaluasi yang diterapkan adalah *accuracy*, *precision*, *recall*, dan *F1-Score* [6]. Evaluasi dilakukan menggunakan data test yang telah di bagi sebelumnya. Perhitungan akurasi menggunakan *confusion matrix* dengan rumus-rumus tersebut:

$$\text{Nilai Accuracy} = \frac{TP + TN}{(TP + FP + TN + FN)} \quad (5)$$

$$\text{Nilai Precision} = \frac{TP}{(TP + FP)} \quad (6)$$

$$\text{Nilai Recall} = \frac{TP}{(TP + FN)} \quad (7)$$

$$\text{Nilai F1-Score} = \frac{\text{precision} \times \text{recall}}{(\text{precision} + \text{recall})} \quad (8)$$

Keterangan :

- (TP) = jumlah data positif yang diprediksi benar
- (TN) = jumlah data yang negatif terprediksi benar
- (FP) = jumlah data yang positif namun terprediksi salah

(FN) = jumlah data berlabel negatif terprediksi salah

3. Hasil dan Diskusi

Penelitian ini bertujuan untuk melakukan analisis sentimen komentar publik, yaitu masyarakat umum ke dalam kategori sentimen positif dan negatif dengan menggunakan Algoritma Multinomial Naive Bayes yang diimplementasikan dengan bahasa pemrograman Python. Sebelum proses analisis sentimen dilakukan ekstraksi fitur menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency) untuk merepresentasikan data teks dalam bentuk numerik. Evaluasi dilakukan dengan melihat performa model berdasarkan metrik akurasi, presisi, recall, dan F1-score.

3.1. Hasil Pre-Processing Text

Berdasarkan data yang didapat dari situs terbuka kaggle, data awal terdiri dari 1517 baris data dengan lima kolom, dan tiga nilai sentimen, yaitu positif negatif netral. Kemudian data dilakukan pre-processing dengan beberapa tahapan, yaitu cleaning, normalisasi, *stopword remover*, tokenisasi, dan stemming dan setiap langkah tahapan tersebut dapat dilihat pada tabel nomor dua.

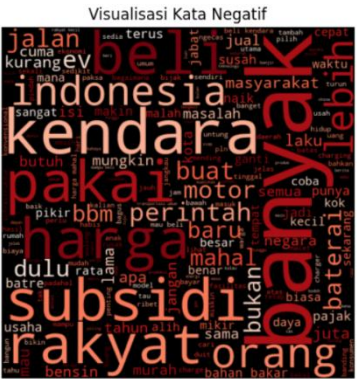
Tabel 1. Hasil Setiap Tahapan Pre-Processing Text

No.	Tahapan	Hasil Setiap Tahapan
1	Data Awal	problem subsidi kualitas diturunin harga dinaikin ...
2	Normalisasi	masalah subsidi kualitas diturunin harga dinaikin ...
3	Stopword Remover	masalah subsidi kualitas diturunin harga dinaikin ...
4	Tokenisasi	[masalah, subsidi, kualitas, diturunin, harga, dinaikin, ..., ...]
5	Stemming	masalah subsidi, kualitas turun harga naik ...

Berdasarkan tabel 2, hasil *pre-processing* dan dengan seleksi data berdasarkan kata kunci berkaitan dengan topik insentif dengan pemilihan data yang memiliki kata kunci terkait, yaitu “subsidi”, “rakyat”, “insentif”, “gratis”, “pln”, dan “program”, didapat data yang berjumlah 368 baris data dan dua kolom, dengan jumlah sentimen negatif, positif dan netral secara berturut-turut, yaitu 228, 107 dan 33 data.



Gambar 1. Word cloud kata positif



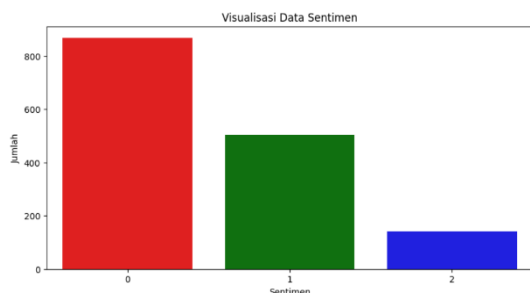
Gambar 2. Word cloud kata negatif



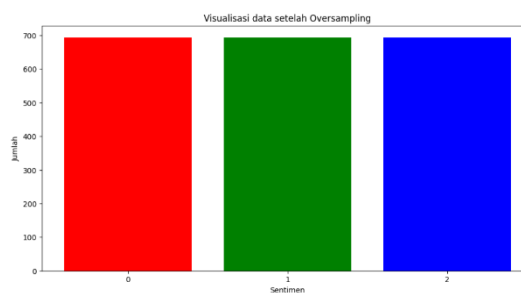
Gambar 3. Word cloud kata netral

Berdasarkan gambar 1, 2 dan 3, dapat dilihat persebaran kata di setiap label sentimen, yaitu sentimen positif, negatif dan netral tersebut. Gambar tersebut merupakan hasil visualisasi dari

data yang sudah melalui tahap *pre-processing* sebelumnya. Dan dapat dilihat kata-kata yang terkandung di setiap label data dimana pada data positif mengandung banyak kata-kata seperti harga, kendra dan subsidi, pada label negatif terdapat kata-kata seperti banyak, subsidi, dan kendra, sedangkan pada data netral terdapat lebih dominan kata subsid, kendra dan orang.



**Gambar 4.** Data sebelum oversampling dengan SMOTE



**Gambar 5.** Data setelah oversampling dengan SMOTE

Berdasarkan Gambar 4 dan 5, dapat dilihat visualisasi data sebelum dan sesudah dilakukannya tahap oversampling untuk meningkatkan jumlah data yang terdapat di label sentimen dengan jumlah data tidak sama atau yang sedikit. Dalam kasus ini adalah dilakukan oversampling untuk data dengan label positif dan netral. Sebelum dilakukan hal ini jumlah dari pesebaran data setiap labelnya yaitu secara berturut-turut untuk label, negatif, positif dan netral, yaitu 228, 107 dan 33, dan dengan oversampling data menjadi sama.

### 3.2. Hasil Implementasi Algoritma *Multinomial Naive Bayes* dan *Ekstraksi Fitur TF-IDF*

Algoritma Multinomial Naive Bayes diimplementasikan dengan membagi data menjadi 80% untuk data pelatihan dan 20% untuk pengujian. Proses ekstraksi fitur yang dilakukan menggunakan metode TF-IDF dan oversampling menggunakan SMOTE. Berdasarkan skenario tersebut diperoleh hasil evaluasi model yang mencakup akurasi, presisi, recall, dan F1-score sebagaimana disajikan dalam tabel 2.

**Tabel 2.** Hasil Evaluasi Model

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
negatif	0.77	0.83	0.80	174
positif	0.71	0.60	0.65	101
netral	0.30	0.32	0.31	28
accuracy			0.71	303
macro avg	0.59	0.58	0.59	303
Weighted avg	0.71	0.71	0.70	303

Berdasarkan hasil evaluasi model yang terdapat pada tabel 2, mode Multinomial Naive Bayes yang diimplementasikan dengan pembagian data latih sebesar 80% dan data uji 20%, menggunakan ekstraksi fitur TF-IDF, serta oversampling dengan SMOTE, dan hyperparameter tuning, yaitu nilai alpha terbaik yaitu 0.1, serta seleksi fitur dengan nilai fitur yang diambil sebanyak 600 fitur, yaitu menunjukkan kinerja terbaik ketika memprediksi data dengan label negatif dengan akurasi 77%, kemudian untuk data berlabel positif sebesar 71%. Namun pada data berlabel netral, model menunjukkan kinerja yang kurang baik yaitu dengan akurasi sekitar 30%, yang memiliki arti masih cukup rendah. Secara Keseluruhan, model miliki akurasi 71%, sehingga dapat dikatakan cukup baik terutama dalam mengenali data dengan label negatif dan negatif.

#### 4. Kesimpulan

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan insentif mobil listrik di Indonesia dengan menggunakan data komentar dari YouTube dengan mengambil data dari beberapa komentar video youtube yang membahas terkait insentif mobil listrik di Indonesia. Dataset ini penulis dapatkan dari situs kaggle dengan kata kunci "mobil listrik". Kemudian Melalui serangkaian tahapan pengumpulan data, pre-processing, data diubah menjadi format numerik menggunakan metode TF-IDF, lalu diklasifikasikan menggunakan algoritma Multinomial Naïve Bayes. Proses oversampling dengan SMOTE dilakukan untuk mengatasi ketidakseimbangan data, seleksi fitur, dan hyperparameter tuning dengan nilai alpha terbaik yaitu 0.1. Hasil evaluasi menunjukkan bahwa model mampu mencapai akurasi sebesar 71%. Model memiliki performa yang lebih baik dalam mengklasifikasikan komentar negatif dibandingkan komentar positif, yang terlihat dari nilai recall dan F1-score yang lebih tinggi pada kelas negatif. Hal ini mengindikasikan bahwa mayoritas opini masyarakat terhadap kebijakan insentif mobil listrik cenderung bersifat kritis. Dengan demikian, hasil penelitian ini dapat menjadi bahan pertimbangan bagi pemerintah dalam mengevaluasi persepsi publik serta merumuskan strategi kebijakan yang lebih tepat sasaran.

#### Daftar Pustaka

- [1] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, p. 406, Apr. 2021, doi: 10.30865/mib.v5i2.2835.
- [2] C. Herdian, A. Kamila, and I. G. Agung Musa Budidarma, "Studi Kasus Feature Engineering Untuk Data Teks: Perbandingan Label Encoding dan One-Hot Encoding Pada Metode Linear Regresi," *Technologia : Jurnal Ilmiah*, vol. 15, no. 1, p. 93, Jan. 2024, doi: 10.31602/tji.v15i1.13457.
- [3] N. Amini *et al.*, "JIP (Jurnal Informatika Polinema) Implementasi Algoritma Genetika Untuk Seleksi Fitur Pada Klasifikasi Genre Musik Menggunakan Metode Random Forest".
- [4] C. H. Yutika, A. Adiwijaya, 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, p. 422, Apr. 2021, doi: 10.30865/mib.v5i2.2845.
- [5] A. Karimah and G. Dwilestari, "Analisis Sentimen Komentar Video Mobil Listrik Di Platform Youtube Dengan Metode Naive Bayes," 2024. [Online]. Available: <https://www.kaggle.com/datasets/billycemerson/anali>
- [6] P. Studi Informatika, F. Matematika dan Ilmu Pengetahuan Alam, J. Raya Kampus UNUD, B. Jimbaran, and K. Selatan, "Analisis Sentimen pada Sengketa Pilpres 2024 dengan Multinomial Naïve Bayes dan Chi-Square I Gede Widnyana a1, AAIN Eka Karyawati a2," *JNATIA*, vol. 3, no. 3, p. 2025.