

# ANALYSIS OPINI PUBLIK TERHADAP KENDARAAN LISTRIK MENGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER (NBC) DENGAN EKSTRAKSI FITUR TF-IDF

[*Analysis Of Public Opinion On Electric Vehicles Using Naïve Bayes Classifier (NBC)  
Algorithm With TF-IDF Feature Extraction*]

L.M. Ryas Amin Akbar<sup>1</sup>, L. Ahmad S. Irfan Akbar<sup>2</sup>, Suthami Ariessaputra<sup>3</sup>

<sup>1,2,3</sup>Jurusan Teknik Elektro Universitas Mataram

<sup>1</sup>[ryasakbar060@gmail.com](mailto:ryasakbar060@gmail.com), <sup>2</sup>[irfan@unram.ac.id](mailto:irfan@unram.ac.id), <sup>3</sup>[suthami@unram.ac.id](mailto:suthami@unram.ac.id)

---

## ABSTRAK

Teknologi telah menciptakan kendaraan listrik, yang dianggap ramah lingkungan dengan motor listrik dan baterai sebagai sumber tenaga. Analisis opini terhadap kendaraan listrik dilakukan dengan data komentar pada Twitter dan YouTube. Dalam melakukan analisis, algoritma *machine learning Naïve Bayes Classifier* digunakan untuk mengklasifikasi pendapat masyarakat menjadi sentimen positif atau negatif. Tujuan penelitian ini untuk menilai persepsi masyarakat, mengevaluasi performa model klasifikasi *Naïve Bayes Classifier* dengan ekstraksi fitur *Term Frequency-Inverse Document Frequency* (Tf-Idf) dan membangun aplikasi analisis opini berbasis *web* untuk klasifikasi data komentar. Pengumpulan data diperoleh dengan teknik *crawling* dan melalui *preprocessing* sebelum di analisis. Model analisis dibangun dengan *Naïve Bayes Classifier* dalam dua konfigurasi yaitu dengan ekstraksi fitur Tf-Idf (NBC TF-IDF) dan tanpa ekstraksi fitur Tf-Idf (NBC). Hasil penelitian menunjukkan model klasifikasi NBC TF-IDF dalam 2 kelas (Positif dan Negatif) mencapai akurasi 80.34%, sedangkan model klasifikasi menggunakan NBC mencapai akurasi 77,49%. Pada skenario 3 kelas (Positif, Netral, dan Negatif), model NBC TF-IDF mencapai akurasi 64,83%, sedangkan model NBC mencapai akurasi 62,44%. Model klasifikasi dengan NBC TF-IDF menunjukkan peningkatan performa model yang lebih baik dalam skenario 2 kelas maupun 3 kelas.

**Kata kunci:** *kendaraan listrik, machine learning, Naïve Bayes Classifier, Term Frequency-Inverse Document Frequency, crawling, preprocessing.*

---

## ABSTRACT

*Technology has created electric vehicles, which are considered environmentally friendly with electric motors and batteries as power sources. Opinion analysis on electric vehicles was conducted with comment data on Twitter and YouTube. In conducting the analysis, the Naïve Bayes Classifier machine learning algorithm is used to classify public opinions into positive or negative sentiments. The purpose of this research is to assess public perception, evaluate the performance of the Naïve Bayes Classifier classification model with Term Frequency-Inverse Document Frequency (Tf-Idf) feature extraction and build a web-based opinion analysis application for comment data classification. Data collection is obtained by crawling techniques and through preprocessing before analysis. The analysis model was built with Naïve Bayes Classifier in two configurations, namely with Tf-Idf feature extraction (NBC TF-IDF) and without Tf-Idf feature extraction (NBC). The results showed that the NBC TF-IDF classification model in 2 classes (Positive and Negative) achieved 80.34% accuracy, while the classification model using NBC achieved 77.49% accuracy. In the 3-class scenario (Positive, Neutral, and Negative), the NBC TF-IDF model achieved an accuracy of 64.83%, while the NBC model achieved an accuracy of 62.44%. The classification model with NBC TF-IDF shows a better improvement in model performance in both 2-class and 3-class scenarios.*

**Keywords:** *Electric vehicle, Machine learning, Naïve Bayes Classifier, Term Frequency-Inverse Document Frequency, crawling, preprocessing.*

---

## 1. PENDAHULUAN

Teknologi modern telah berkembang begitu pesat dengan membawa banyak perubahan di berbagai bidang, salah satunya bidang transportasi telah menciptakan inovasi baru yaitu kendaraan listrik. Kendaraan listrik adalah kendaraan yang menggunakan motor listrik sebagai sumber tenaga dan baterai sebagai penyimpan energi serta berpotensi sebagai solusi ramah lingkungan. Dengan memanfaatkan data Twitter dan Youtube maka dilakukan analisis opini publik terhadap kendaraan listrik dalam menganalisis pendapat, perasaan, penilaian, sikap dan emosi berdasarkan bahasa tulis atau teks komentar (Fatihin, 2022). Topik tersebut cukup menarik untuk diteliti guna menganalisis respon masyarakat berdasarkan data teks komentar dengan memanfaatkan algoritma machine learning Naïve Bayes Classifier sebagai metode pengklasifikasian untuk memprediksi probabilitas pendapat masyarakat kedalam sentimen positif atau negatif, sesuai dengan perhitungan probabilitas.

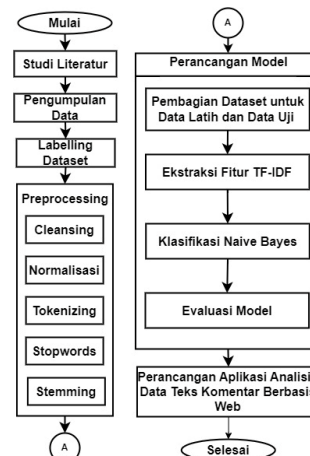
Beberapa penelitian telah dilakukan untuk analisis opini berdasarkan data teks komentar dengan algoritma Naïve Bayes Classifier diantaranya yaitu, Penelitian oleh (Irawansyah dkk, 2022) melakukan analisis sentimen terhadap program kampus merdeka pada twitter menggunakan Naïve Bayes Classifier, dimana teknik labelling data dalam penelitian ini dilakukan secara otomatis dengan metode TextBlob. Namun, kekurangan labelling data dengan metode TextBlob ini tidak support penuh terhadap data teks berbahasa Indonesia. Kemudian, penelitian oleh (Toy dkk, 2021) melakukan penelitian menggunakan metode Naïve Bayes dengan Relevance Frequency Feature Selection terhadap kebijakan New Normal. Namun, kekurangan dari penelitian ini yaitu penggunaan Metode Naïve Bayes dan Relevance Frequency Feature Selection menghasilkan akurasi yang masih terbilang rendah dikarenakan datah latih yang digunakan terlalu sedikit dan kurang beragam. Selanjutnya, penelitian oleh (Riyadi dkk, 2022) melakukan pengukuran sentimen sosial terhadap teknologi kendaraan listrik untuk mengetahui sentimen publik terhadap kendaraan listrik melalui komentar di twitter menggunakan Naïve Bayes, K-NN dan

Decision Tree. Pada penelitiannya visualisasi hasil dan analisis data dilakukan secara manual dengan tools rapidminer.

Berdasarkan uraian diatas, maka dalam penelitian ini penulis melakukan Analisis Opini Publik Terhadap Kendaraan Listrik Menggunakan Algoritma Naïve Bayes Classifier (NBC) dengan Ekstraksi Fitur Term Frequency-Inverse Document Frequency (TF-IDF). Pada penelitian ini digunakan metode yang berbeda dalam proses labelling data, dimana dilakukan labelling manual (Manual Classification) dengan bantuan pakar bahasa Indonesia yaitu Ibu Dra. Syamsinas Jafar, M. Hum yang merupakan dosen Bahasa Indonesia FKIP Universitas Mataram. Kemudian, penelitian ini menggunakan ekstraksi fitur dengan Tf-Idf untuk melakukan perhitungan nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) pada setiap fitur kata di setiap dokumen dataset. Pada penelitian ini juga dilakukan implementasi model mechine learning kedalam sistem aplikasi analisis opini berdasarkan data teks komentar berbasis web menggunakan framework streamlit, sehingga proses klasifikasi, visualisasi dan evaluasi data analisis ditampilkan pada sistem aplikasi berbasis web.

## 2. METODE PENELITIAN

Penelitian ini dilakukan berdasarkan alur penelitian pengerjaan tugas akhir yang terdapat pada Gambar 1 Untuk memastikan hasil yang dicapai tidak menyimpang dan sesuai dengan tujuan penelitian.



Gambar 1. Alur Penelitian

Tahap pertama adalah Studi literatur untuk memahami konsep dan dasar teori yang akan menjadi pembanding dan pendukung dalam penelitian berjudul "Analysis Opini Publik Terhadap Kendaraan Listrik Menggunakan Algoritma Naïve Bayes Classifier (Nbc) Dengan Ekstraksi Fitur Term Frequency-Inverse Document Frequency (Tf-Idf)".

Tahap kedua adalah pengumpulan *data*, dalam penelitian ini data yang digunakan merupakan data komentar dari Twitter dan Youtube yang di kumpulkan dengan teknik crawling menggunakan rapidminer dan layanan google youtube data API v3 serta library `googleapiclient.discovery` yang memungkinkan untuk berinteraksi dengan youtube Data API dan mengambil data komentar dari video youtube. Data komentar yang diambil adalah data komentar dari video youtube yang membahas mengenai kendaraan listrik dan perkembangan teknologi tersebut. Variabel meta data hasil crawling yang akan diambil adalah username, teks komentar serta tanggal dan waktu komentar.

Tahap ketiga adalah labelling dataset yang dilakukan untuk memberi label kelas sentimen pada data teks komentar yang telah didapatkan menjadi kelas positif, negatif dan netral. Pada proses ini dilakukan pelabelan data ulasan secara manual (manual classification) dengan bantuan dari pakar bahasa yaitu Ibu Dra. Syamsinas Jafar, M. Hum yang merupakan Dosen Bahasa Indonesia FKIP Universitas Matram.

Tahap keempat adalah preprocessing, tahap ini dilakukan untuk membersihkan dataset hasil crawling menjadi data yang bersih dan siap dilakukan pengolahan dalam analisis data. Terdapat beberapa tahap preprocessing yang dilakukan diantaranya yaitu: cleansing, untuk menghilangkan karakter-karakter spesial yang membuat data menjadi tidak efektif; normalisasi, untuk mengubah kata slang atau tidak baku menjadi bahasa yang baku sesuai Kamus Besar Bahasa Indonesia (KBBI); tokenizing, untuk memecah teks atau kalimat menjadi unit-unit yang lebih kecil, yang disebut token; stopwords, untuk menghapus kata-kata yang kurang bermakna atau tidak memiliki arti seperti kata: "dan", "atau", "yang", "di", dan sebagainya; dan stemming, untuk mengubah kata ke dalam bentuk kata dasar dengan

menghapus kata imbuhan di depan maupun imbuhan di belakang kata.

Tahap kelima adalah pembagian dataset, dalam tahapan ini dataset yang telah melalui preprocessing dibagi menjadi data latih dan data uji. Data latih digunakan untuk ekstraksi fitur dengan TF-IDF serta melatih model klasifikasi Naïve Bayes sehingga model dapat belajar melalui data dan dapat mengenali pola dari data latih tersebut. Sedangkan data uji akan digunakan untuk mengukur kinerja model agar dapat melihat persentase algoritma klasifikasi yang berhasil mengklasifikasikan dengan benar.

Tahap keenam adalah ekstraksi fitur, dalam tahap ini dilakukan perubahan fitur kata menjadi sebuah representasi angka (numerik) dengan melihat kemunculan fitur kata terlebih dahulu dengan metode `CountVectorizer`, lalu dilanjutkan dengan mencari pembobotan kata menggunakan metode Tf-Idf yang akan melakukan perhitungan nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) pada setiap fitur kata di setiap dokumen dalam dataset.

Tahap ketujuh adalah klasifikasi, dalam tahap ini Naïve Bayes akan melakukan prediksi kelas positif atau negatif dengan menguji data serta mempelajari pengetahuan yang terkandung dalam data latih. Pada data latih terdapat kelas positif dan negatif, Naïve Bayes akan mempelajari karakteristik kata-kata yang terdapat pada masing-masing kelas.

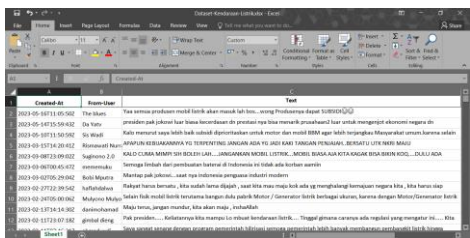
Tahap kedelapan adalah evaluasi model, tahapan ini dilakukan untuk mengetahui kinerja model klasifikasi. Evaluasi model dilakukan dengan cara melihat tingkat akurasi, presisi, recall dan f1-score yang mengacu pada hasil tabel *confusion matrix* yang dihasilkan.

Tahap terakhir adalah perancangan aplikasi, dalam merancang sistem aplikasi yang dapat digunakan untuk analisis opini berdasarkan teks komentar dengan mengimplementasikan metode Naïve Bayes Classifier dibuatlah sebuah diagram alir untuk membantu dalam proses memahami alur ketika sistem berlangsung. Selain itu pada tahap ini juga design user interface akan dibuat sebelum dilakukan implementasi kedalam web dengan menggunakan framework streamlit dari pemrograman python.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Pengumpulan Data

Pengumpulan data dalam penelitian ini didapatkan melalui teknik crawling data komentar pada *twitter* dan *youtube* yang membahas mengenai kendaraan listrik dengan rentang waktu data yang diambil mulai dari 15 April 2021 – 25 Mei 2023. Kemudian, dataset hasil crawling disimpan dengan format excel dan dapat diakses pada link github berikut: (<https://github.com/ryasakbar060/Dataset-Komentar-Kendaraan-Listrik>). Dataset hasil crawling dalam format excel dapat dilihat pada Gambar 2.



Gambar 2. File excel dataset hasil crawling komentar twitter dan youtube

#### 3.3. Labelling Data

Penentuan kelas sentimen (labelling data) dilakukan secara manual (manual classification) dengan bantuan dari pakar bahasa yaitu Ibu Dra. Syamsinas Jafar, M. Hum yang merupakan Dosen Bahasa Indonesia FKIP Universitas Mataram. Dalam penelitian ini dilakukan penentuan sentimen secara manual oleh penulis terlebih dahulu dengan menentukan tiga kelas yaitu Positif, Negatif dan Netral. Setelah dilakukan pelabelan data secara manual oleh penulis, data hasil labelling akan diteruskan dan di koreksi kembali oleh pakar.

Pada Tabel 3.1 berisi detail jumlah data hasil labelling awal dataset pada masing masing kelas sentimen.

Tabel 3.1 Hasil labelling awal dataset

Positif	Negatif	Netral	Total
1972	1620	678	4270

Pada Tabel 3.2 berisi detail jumlah data hasil labelling setelah kelas Netral dihapus dalam dataset.

Tabel 3.2 Detail labelling data Positif dan Negatif

Positif	Negatif	Total
1972	1620	4270

Kemudian, data pada Tabel 3.2 dibersihkan pada tahap preprocessing sehingga menghasilkan data bersih sebesar 3509 yang digunakan untuk membangun model machine learning analisis sentimen menggunakan Naive Bayes Classifier dengan ekstraksi fitur Tf-Idf dan menghasilkan detail data bersih hasil labelling seperti pada Tabel 3.3.

Tabel 3.3 Detail labelling data setelah preprocessing

Positif	Negatif	Total
1972	1620	4270

#### 3.4. Preprocessing

Preprocessing yang dilakukan terdiri atas beberapa tahapan diantaranya yaitu cleansing, normalisasi, tokenizing, stopwords dan stemming.

##### a) Cleansing

```
import re
import emoji

def cleansing(Text):
    Text = re.sub("RT", "", Text) # remove RT
    Text = Text.replace("<br>", " ") # mengubah <br> menjadi spasi
    Text = ' '.join(re.sub("([@#][A-Za-z0-9]+)|(\w+:\w+\/\w+)", "", Text).split()) # Menghapus mention, hashtag, dll
    Text = Text.lower() # Mengubah menjadi huruf kecil
    Text = re.sub("([a-z0-9]+)", "", Text)
    Text = emoji.demojize(Text) # Menghilangkan emoji
    Text = re.sub("([a-zA-Z_]+)", "", Text)
    Text = re.sub("([\w\W])", "", Text) # remove tanda baca
    Text = re.sub("([d+]", "", Text) # remove angka
    Text = Text.replace("http://", "").replace("https://", "") # remove URL
    Text = Text.replace("\t", " ") .replace("\n", " ")
    Text = Text.replace("\u", " ") .replace("\u", " ") # Menghapus karakter escape seperti '\t', '\n', '\u'
    return Text
```

Gambar 3. Tahapan Cleansing

Gambar 3 adalah kode untuk melakukan Cleansing guna membersihkan data teks dari hal yang membuat data menjadi tidak efektif seperti hashtag, mention username, URL, tanda baca, double space, emoticon, bilangan angka, simbol dan case folding.

##### b) Normalisasi

```
def slang_normalization(text):
    df_slang = pd.read_excel("normalisasi.xlsx")
    slang_dict = dict(zip(df_slang["original"], df_slang["replacement"]))
    text = ' '.join([slang_dict[word] if word in slang_dict else word for word in text.split()])
    return text

df['Normalisasi'] = df['Cleansing'].apply(slang_normalization)
df.head()
```

Gambar 4. Tahapan Normalisasi

Gambar 4 adalah kode untuk melakukan Normalisasi yang bertujuan mengubah kata tidak baku (slang words) dalam data teks menjadi bahasa yang baku sesuai Kamus Besar Bahasa Indonesia (KBBI) dan acuan file kamus normalisasi (normalisasi.xlsx)

c) Tokenizing

```
from nltk.tokenize import word_tokenize

def tokenization(Text):
    tokens = word_tokenize(Text)
    return tokens

df['Tokenize'] = df['Normalisasi'].apply(tokenization)
df.head()
```

Gambar 5. Tahapan Tokenizing

Gambar 5 adalah kode untuk melakukan Tokenizing yang bertujuan memecah teks atau kalimat menjadi unit-unit yang lebih kecil, yang disebut token sehingga memungkinkan mengubah teks menjadi lebih terstruktur, kemudian dapat diolah lebih lanjut untuk memahami sentimen yang terkandung di dalamnya.

d) Stopwords

```
from nltk.corpus import stopwords

def remove_stopwords(tokens):
    list_stopwords = nltk.corpus.stopwords.words('indonesian')
    list_stopwords.extend(['yg', 'dg', 'dgn', 'ny', 'di', 'tu',
        'klo', 'kalo', 'amp', 'kian', 'bikin', 'bilang', 'gak', 'ga',
        'krn', 'nya', 'nih', 'sih', 'si', 'sau', 'tdk', 'suh', 'utk',
        'ya', 'jd', 'jgn', 'sdh', 'aja', 'n', 't', 'p', 'ak',
        'nyg', 'haha', 'pen', 'tu', 'man', 'loh', 'rt', 'samp', 'yah',
        'ra', 'des', 'di', 'lme', 'ge', 'eeehhh'])
    txt_stopword = pd.read_csv('stopwords.txt',
        names=['stopwords'], header=None)
    list_stopwords.extend(txt_stopword['stopwords'][0].split(' '))
    list_stopwords = set(list_stopwords)
    tokens = [word for word in tokens if not word in
        list_stopwords]
    return tokens

df['StopWord'] = df['Tokenize'].apply(remove_stopwords)
df.head()
```

Gambar 6. Tahapan Stopwords

Gambar 6 adalah kode untuk melakukan Stopwords yang bertujuan menghilangkan kata-kata yang tidak memiliki makna khusus dalam analisis teks agar membuat data menjadi lebih bersih dan memfokuskan perhatian pada kata-kata yang lebih relevan dan informatif.

e) Stemming

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

def stemming(Text):
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    Text = [stemmer.stem(word) for word in Text]
    return Text

df['Stemming'] = df['StopWord'].apply(stemming)
df.head()
```

Gambar 7. Tahapan Stemming

Gambar 7 adalah kode untuk melakukan Stemming bertujuan mengubah kata ke dalam bentuk kata dasarnya dengan menghapus imbuhan di depan maupun imbuhan di belakang kata atau mereduksi kata dalam teks menjadi bentuk dasar, sehingga kata-kata yang memiliki arti yang sama atau terkait akan dikelompokkan menjadi satu bentuk.

### 3.5. Pembagian Dataset

Berikut pada Tabel 3.4 beberapa perbandingan data latih dan data uji yang dilakukan pada penelitian ini beserta akurasi yang dihasilkan.

Tabel 3.4 Perbandingan penggunaan data latih dan data uji

Perbandingan	Detail Data		Akurasi	Total Data
	Data Latih	Data Uji		
90 : 10	3158	351	81%	3509
80 : 10	2807	702	78%	
70 : 30	2456	1053	77%	
60 : 40	2105	1404	76%	

Tabel 3.4 diatas menunjukkan bahwa dari keempat perbandingan yang digunakan didapatkan perbandingan 90:10 menghasilkan performa atau nilai akurasi terbaik sehingga perbandingan ini dipilih untuk digunakan dalam melakukan analisis opini publik terhadap kendaraan listrik menggunakan algoritma naive bayes classifier dan ekstraksi fitur dengan tf-idf.

### 3.6. Ekstraksi Fitur

Pada tahap ekstraksi fitur ini bertujuan untuk mengubah dataset kedalam representasi numerik menggunakan library python Scikit Learn dengan metode CountVectorizer untuk melihat berapa kali fitur kata muncul dalam suatu document pada dataset, kemudian dilakukan pembobotan kata dengan menggunakan metode TF-IDF yang juga merupakan metode dalam Scikit Learn, dimana tujuan dilakukannya ekstraksi fitur dengan TF-IDF adalah untuk mengidentifikasi kata-kata yang paling relevan dan berpengaruh dalam suatu dokumen berdasarkan nilai pembobotan kata yang didapatkan. Berikut pada Tabel 3.5 dapat dilihat untuk sampel daftar nilai pembobotan kata yang memiliki nilai rata-rata tertinggi.

Tabel 3.5. Sampel daftar nilai pembobotan kata Tf-Idf

Kata	Nilai Tf-Idf
listrik	0,085291
kendara	0,073589
mobil	0,056195
indonesia	0,036669
subsidi	0,034352
baterai	0,022980
perintah	0,022494
motor	0,022307
beli	0,021205
pakai	0,020124

### 3.7. Klasifikasi Naïve Bayes

Untuk melakukan klasifikasi Naïve Bayes, dataset komentar pada penelitian ini harus melalui tahap pembagian dataset terlebih dahulu menjadi data latih dan data uji. Seperti yang telah dijelaskan sebelumnya perbandingan yang digunakan dalam klasifikasi yaitu 90:10. Berikut proses melatih dan menguji model Naïve Bayes Classifier dalam analisis sentimen opini publik terhadap kendaraan listrik.

#### a) Melatih model Naïve Bayes

```
from nbcc import MultinomialNB
x_train = train_tfidf
y_train = train_labels
modelNBC = MultinomialNB()
modelNBC.fit(x_train, y_train)
```

Gambar 8. Tahap pelatihan model Naïve Bayes

Gambar 8 adalah kode untuk implementasi proses pelatihan sebuah model klasifikasi Naïve Bayes yang bertujuan untuk melatih model dalam mempelajari dan mengenali pola dalam data latih.

#### b) Menguji Model Naïve Bayes

```
x_test = test_tfidf
y_test = test_labels
predictNBC = modelNBC.predict(x_test)
predictNBC
```

Gambar 9. Tahap pengujian model Naïve Bayes

Gambar 9 adalah kode untuk implementasi model klasifikasi Naive Bayes yang telah dilatih digunakan untuk memprediksi kelas pada 351 data uji dengan hasil klasifikasi 192 sentimen positif dan 159 negatif. Sampel data hasil klasifikasi Naive Bayes berdasarkan data uji dapat dilihat pada tabel 3.6.

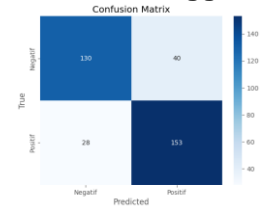
Tabel 3.6 Hasil klasifikasi model Naïve Bayes

Comment	Label	Classification
pabrik mobil konvensional mobil konvensional untung	Positif	Negatif
energi baru tinggal mau perintah sungai indonesia ribu sumber tenaga matahari angin berimpah bukti maju mafia koruptor cokor perintah	Negatif	Negatif
perintah resmi terbit atur beri bantu beli kendara motor listrik bas baterai	Positif	Positif
sedih negara asai pihak guna kendara listrik guna kri subsidi kendara listrik triliun rupiah gampang gelontor gilir nambah gerbong bekas ditolakin	Negatif	Negatif
mobil buang mesin ganti dinamo batraae charger solar sell	Negatif	Negatif

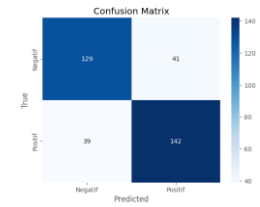
### 3.8. Evaluasi Model

Evaluasi model bertujuan untuk mendapatkan performa model machine learning yang telah dibuat dalam melakukan analisis sentimen. Dalam melakukan evaluasi model, akan dilihat nilai-nilai seperti akurasi, presisi, recall, dan F1-Score. Untuk mendapatkan nilai-nilai tersebut, maka dilakukan pengujian confusion matrix untuk evaluasi hasil dari prediksi yang dilakukan oleh model. Berikut adalah hasil confusion matrix dari model Naïve Bayes dengan ekstraksi fitur Tf-Idf dan tanpa ekstraksi fitur Tf-Idf serta membandingkan penggunaan 2 kelas (Positif dan Negatif) dengan 3 kelas (Positif, Netral dan Negatif).

#### 3.8.1 Evaluasi Model menggunakan 2 Kelas



Gambar 10. Confusion Matrix Naïve Bayes TF-IDF



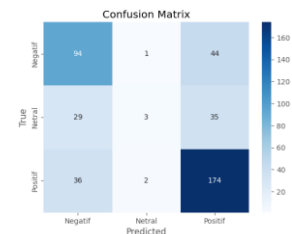
Gambar 11. Confusion Matrix Naïve Bayes tanpa TF-IDF

Berdasarkan kedua hasil confusion matrix diatas didapatkan perbandingan performa seperti pada Tabel 3.7 berikut.

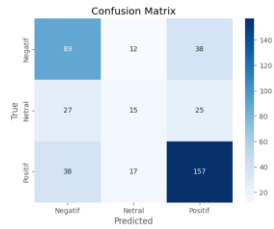
Tabel 3.7 Perbandingan performa model 2 kelas

Model	Akurasi	Presisi		Recall		F1-Score	
		Positif	Negatif	Positif	Negatif	Positif	Negatif
NBC TF-IDF	80,62%	0,79	0,82	0,85	0,76	0,82	0,79
NBC	77,20%	0,78	0,77	0,78	0,76	0,78	0,76

#### 3.8.1 Evaluasi Model menggunakan 3 Kelas



Gambar 12. Confusion Matrix Naïve Bayes TF-IDF



Gambar 13. Confusion Matrix Naïve Bayes tanpa TF-IDF

Berdasarkan kedua hasil confusion matrix diatas didapatkan perbandingan performa seperti pada Tabel 3.8 berikut.

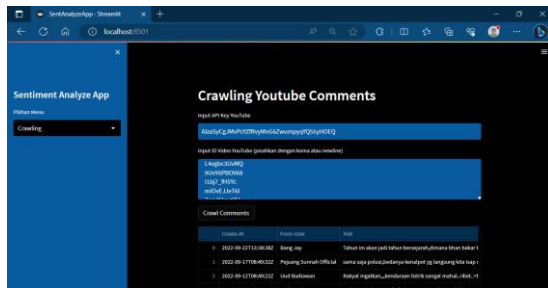
Tabel 3.8 Perbandingan performa model 3 kelas

Model	Akurasi	Presisi			Recall			F1-Score		
		Positif	Negatif	Netral	Positif	Negatif	Netral	Positif	Negatif	Netral
NBC TF-IDF	64,83%	0,69	0,59	0,50	0,82	0,68	0,04	0,75	0,63	0,08
NBC	62,44%	0,71	0,58	0,34	0,74	0,64	0,22	0,73	0,62	0,27

### 3.9. Perancangan Aplikasi

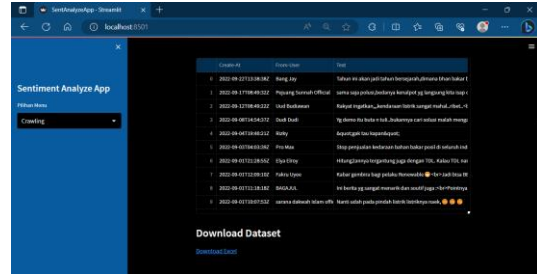
Setelah melalui tahapan perancangan model analisis opini publik terhadap kendaraan listrik menggunakan algoritma Naïve Bayes Classifier. Selanjutnya akan dilakukan implementasi model klasifikasi machine learning kedalam sistem berbasis website yang dibangun menggunakan framework streamlit. Website ini dapat melakukan klasifikasi dataset komentar untuk melihat sentimen masyarakat terhadap kendaraan listrik dengan algoritma Naïve Bayes Classifier. Berikut penyajian hasil perancangan aplikasi dengan beberapa fitur yang disajikan.

#### 3.9.1 Fitur Crawling



Gambar 14. Tampilan website sentimen analisis pada fitur Crawling

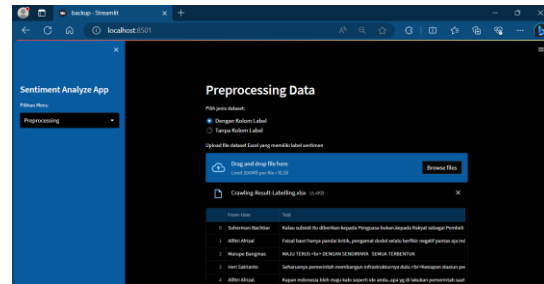
Gambar 14 adalah tampilan pada fitur Crawling yang dimana pada fitur ini sistem akan melakukan pengambilan data teks komentar pada video youtube dengan menginputkan API Key youtube yang telah dibuat melalui google cloud dan ID Video youtube yang ingin diambil komentarnya.



Gambar 15. Tampilan hasil crawling data teks komentar

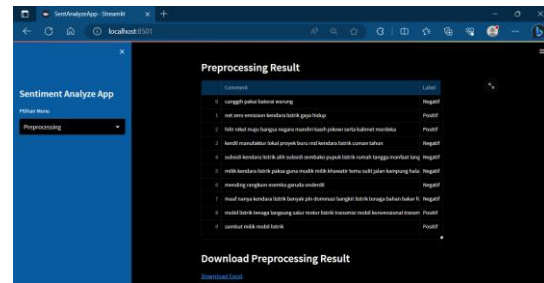
Gambar 15 adalah tampilan hasil data teks komentar setelah dilakukan crawling dengan API Key dan ID Video youtube yang diinputkan dan menghasilkan tabel data dengan variable yang diambil adalah tanggal serta waktu komentar (Create-At), From-User dan teks komentar (Text). Dimana, data hasil crawling dapat di download dengan format excel.

#### 3.9.2 Fitur Preprocessing



Gambar 16. Tampilan website sentimen analisis pada fitur Preprocessing

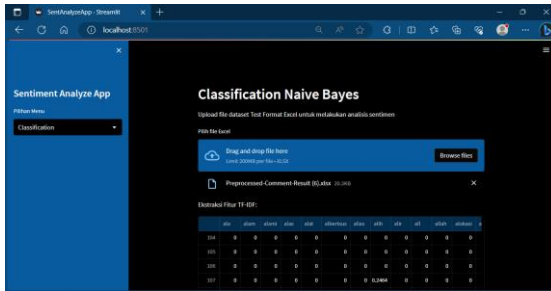
Gambar 16 adalah tampilan pada fitur Preprocessing yang dimana pada fitur ini sistem akan melakukan tahap preprocessing untuk membersihkan data sehingga dapat dilakukan pengklasifikasian pada tahap selanjutnya.



Gambar 17. Tampilan hasil dataset setelah preprocessing pada fitur Preprocessing

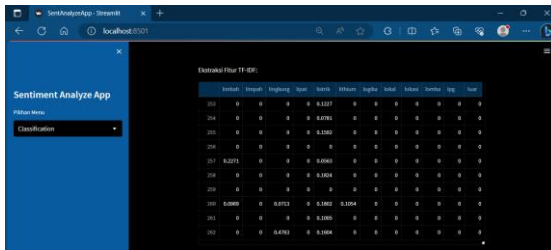
Gambar 16 adalah tampilan hasil data setelah dilakukan preprocessing atau pembersihan dataset oleh sistem, dimana menampilkan tabel data yang sudah bersih dan dapat di download dalam format excel.

### 3.9.3 Fitur Classification



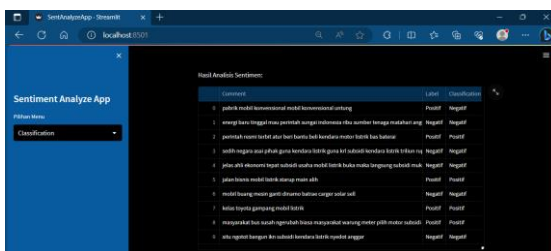
Gambar 18. Tampilan website sentimen analisis pada fitur Classification

Gambar 18 adalah tampilan fitur Classification, dimana pada fitur ini sistem akan melakukan klasifikasi dataset uji dengan format excel yang diinputkan untuk dilakukan analisis opini publik terhadap kendaraan listrik menggunakan algoritma Naïve Bayes Classifier.



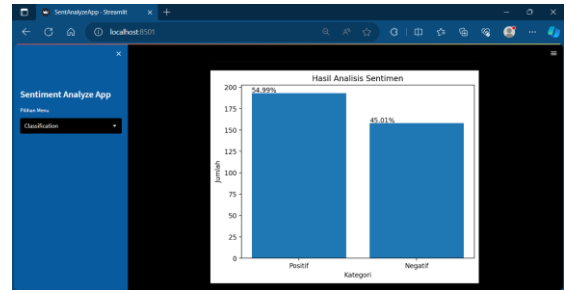
Gambar 19. Tampilan ekstraksi fitur Tf-Idf pada fitur Classification

Gambar 19 adalah tampilan fitur Classification yang menampilkan proses ekstraksi fitur yang dilakukan sistem dengan metode Tf-Idf untuk mencari pembobotan fitur kata.



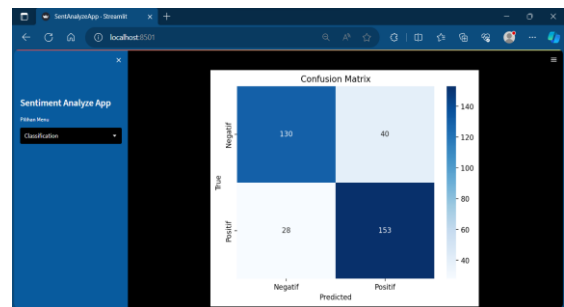
Gambar 20. Tampilan data hasil klasifikasi sentimen pada fitur Classification

Gambar 20 adalah tampilan fitur Classification yang menampilkan data hasil klasifikasi dari data uji yang dilakukan oleh sistem.



Gambar 21. Tampilan grafik persentase klasifikasi sentimen pada fitur Classification

Gambar 21 adalah tampilan fitur Classification yang menampilkan data hasil klasifikasi dari data uji yang dilakukan oleh sistem dengan persentase grafik dari kelas positif dan negatif.



Gambar 22. Tampilan grafik confusion matrix pada fitur Classification

Gambar 22 adalah tampilan fitur Classification yang menampilkan data hasil klasifikasi dari data uji yang dilakukan oleh sistem dengan grafik confusion matrix.

	precision	recall	f1 score	support
Negatif	0.8000	0.7000	0.7500	170,000
Positif	0.7900	0.8000	0.8000	181,000
accuracy	0.8100	0.8100	0.8100	0.8100
macro avg	0.8100	0.8100	0.8100	351,000
weighted avg	0.8100	0.8100	0.8100	351,000

Gambar 23. Tampilan classification report pada fitur Classification

Gambar 23 adalah tampilan fitur Classification yang menampilkan tabel classification report untuk melihat performa model klasifikasi Naïve bayes untuk nilai akurasi, presisi, recall dan f1-score.



#### 4. KESIMPULAN

1. Dari 3509 data teks komentar yang digunakan melalui hasil crawling pada twitter dan youtube yang telah melalui preprocessing menunjukkan bahwa, dengan pembagian data pada rasio 90:10. Model berhasil mengklasifikasikan sejumlah 192 sentimen Positif dan 159 Negatif dari 351 data uji.
2. Model klasifikasi analisis opini publik terhadap kendaraan listrik menggunakan algoritma Naïve Bayes Classifier dengan ekstraksi fitur Tf-Idf pada penggunaan 2 kelas (Positif dan Negatif) menghasilkan nilai akurasi 81%, presisi 79%, recall 85% dan f1-score 82%. Sedangkan, model klasifikasi tanpa ekstraksi fitur Tf-Idf mendapatkan akurasi 77%, presisi 78%, recall 78% dan f1-score 78%.
3. Model klasifikasi analisis opini publik terhadap kendaraan listrik menggunakan algoritma Naïve Bayes Classifier pada 3 kelas (Positif, Netral dan Negatif) Naïve Bayes Classifier dengan ekstraksi fitur Tf-Idf mendapat akurasi 64%, presisi 69%, recall 82% dan f1-score 75%. Sedangkan, klasifikasi tanpa ekstraksi fitur Tf-Idf mendapat akurasi 62%, presisi 71%, recall 74% dan f1-score 73%.
4. Penerapan model klasifikasi analisis menggunakan Naïve Bayes Classifier dilakukan dengan pengintegrasian model klasifikasi kedalam aplikasi berbasis web yang dibangun menggunakan framework streamlit dengan menerapkan beberapa proses dalam perancangan model yang telah dibuat sebelumnya menjadi beberapa fitur seperti crawling, preprocessing, klasifikasi, visualisasi, dan evaluasi performa model.

#### DAFTAR PUSTAKA

- Agustian, A., Tukino, & Nurapriani, F. (2022). Penerapan Analisis Sentimen Dan Naïve Bayes Terhadap Opini Penggunaan Kendaraan Listrik Di Twitter. *Jurnal Tika Fakultas Ilmu Komputer Universitas Almuslim*, 7(3), 243-249.
- Anwar, M., & Permana, D. (2023). Analisis Sentimen Masyarakat Indonesia Terhadap Produk Kendaraan Listrik Menggunakan VADER. *Jurnal Teknik Informatika dan Sistem Informasi*, 10(1), 783-792.
- Fatihin, A. (2022). Analisis Sentimen Terhadap Ulasan Aplikasi Mobile Menggunakan Metode Support Vector Machine [Skripsi, Universitas Islam Negeri Syarif Hidayatullah]. Repositori Universitas Islam Negeri Syarif Hidayatullah.
- Imron, A. (2019). Analisis Sentimen Terhadap Wisata Di Kabupaten Rembang Menggunakan Metode Naïve Bayes Classifier [Skripsi, Universitas Islam Indonesia]. Repositori Universitas Islam Indonesia.
- Irawansyah, R. S. (2022). Analisis Sentimen Terhadap Program Merdeka Belajar-Kampus Merdeka Pada Twitter Menggunakan Algoritma Naive Bayes Classifier (NBC) [Skripsi, Universitas Mataram]. Repositori Universitas Mataram.
- Khomsah, S., & Aribowo, A. S. (2020). Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia. *Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)*, 4(4), 648-654.
- Pangestu, D. A. (2020). Analisis Sentimen Terhadap Opini Publik Tentang Kesehatan Mental Selama Pandemi Covid-19 Di Media Sosial Twitter Menggunakan Naïve Bayes Classifier Dan Support Vector Machine [Skripsi, Universitas Islam Indonesia]. Repositori Universitas Islam Indonesia.
- Pratama, P., Murdiansyah, D., & Lhaksana, K. (2023). Analisis Sentimen Kendaraan Listrik Pada Media Sosial Twitter Menggunakan Algoritma Logistic Regression dan Principal Component Analysis. *Jurnal Media Informatika Budidarma*, 7(1), 592-535.
- Qinanda, M. D., Nilogiri, A., & Timur, T. W. (2022). Sentimen Pada Komentar Youtube Tentang Pencegahan Dan Penanganan Kekerasan Seksual Pada Permendikbud Berbasis Naive Bayes Dan Support Vector Machine. *Jurnal Sistem dan Teknologi Informasi (JUSTINDO)*, 7(2), 114-121.
- Riyadi, A. F., Rahman, F. R., Pratama, M. A., Khafidli, M. K., & H. P. (2022). Pengukuran Sentimen Sosial Terhadap Teknologi Kendaraan Listrik: Bukti Empiris di Indonesia. *Jurnal Manajemen Sistem Informasi Dan Teknologi*, 12(2), 141-149.
- Santoso, A., Nugroho, A., & Sunge, A. S. (2022). Analisis Sentimen Tentang Mobil Listrik Dengan Metode Support Vector Machine Dan Feature Selection Particle Swarm Optimization. *Jurnal Of Practical Computer Science*, 2(1), 24-31.
- Sasmita, A., Pradnyana, G. A., & Divayana, D. G. (2022). Sistem Analisis Sentimen Untuk Evaluasi Kinerja Dosen Dengan Metode Naive Bayes. *Jurnal Sains dan Teknologi*, 11(2), 451-462.
- Siswandi, A. (2019). Implementasi Data Mining Dengan Metode Klasifikasi Naive Bayes Untuk Memprediksi Stok Bahan Jadi. *Jurnal Teknologi Pelita Bangsa*, 10(2), 1-10.
- Toy, K. V., Sari, Y. A., & Cholissodin, I. (2021). Analisis Sentimen Twitter Menggunakan Metode Naive Bayes Dengan Relevance Frequency Feature Selection (Studi Kasus: Opini Masyarakat Mengenai Kebijakan New Normal). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(11), 5068-5074.