

Analisis Sentimen Terhadap Kendaraan Listrik Pada Platform *Twitter* Menggunakan Metode *Naive Bayes*

Feni Nuriy Alin ^{1,*}, Moch Hafid Totohendarto ², Muhammad Rafi Muttaqin ³

^{1,2,3} Teknik Informatika; Sekolah Tinggi Teknologi Wastukencana; Jl. Cikopak NO. 53, Sadang, Purwakarta 41151, Telp. (0264) 214952; e-mail: feninuriy48@wastukencana.ac.id, mhafid@wastukencana.ac.id, rafi@wastukencana.ac.id.

* Korespondensi: e-mail: feninuriy48@wastukencana.ac.id

Diterima: 28 Juli 2023 ; Review: 15 November 2023; Disetujui: 1 Desember 2023

Cara sitasi: Alin FN, Totohendarto MH, Muttaqin MR, 2023. Analisis Sentimen Terhadap Kendaraan Listrik Pada Platform *Twitter* Menggunakan Metode *Naive Bayes*. Informatics for Educators and Professionals : Journal of Informatics. Vol 8 (2): 96 - 107.

Abstrak: Kehadiran Kendaraan Listrik menimbulkan pro dan kontra dikalangan masyarakat seperti yang ramai dibicarakan di media sosial *twitter*. Sebagian kalangan masyarakat berpendapat kendaraan listrik di klaim efektif menjadi solusi *deficit* migas tetapi sebagiannya lagi berpendapat bahwa unuk menuju kendaraan listrik persiapan yang matang, khususnya dari segi infrastruktur. Melihat banyaknya opini yang ada di *Twitter*, penting untuk melakukan klasifikasi sentimen guna memudahkan identifikasi kecenderungan opini terhadap kendaraan listrik, apakah cenderung positif, negatif, atau netral. Dalam penelitian ini peneliti menggunakan *platform twitter* yang terdiri dari tahapan *crawling data*, *cleaning*, *preprocessing* (*case folding*, *tokenizing*, *stopword removal*), *labelling*. *Naive Bayes* dipilih sebagai algoritma yang digunakan karena memiliki nilai probabilitas atau peluang tertinggi dalam klasifikasi sentimen. untuk pengujian data menggunakan *confusion matrix* bahasa pemrograman *python* menggunakan *Google Colaboratory*. hasil penelitian ini terdapat analisis sentimen mengenai kendaraan listrik dengan jumlah data 1.301, positif sebanyak 940 data, negatif 120 data, dan netral 241 data. Berdasarkan hasil pengujian *confusion matrix* bahasa pemrograman *Python* menggunakan *Google Colaboratory* didapatkan hasil akurasi 75% dengan nilai *precision* positif 81%, negatif 15%, Netral 30% dan tingkat keberhasilan *recal* positif 75%. Nilai ini menunjukkan bahwa klasifikasi algoritma *naive bayes* dinilai cukup baik dalam pemrosesan dokumen *twitter* dikarenakan persentase akurasinya sebesar 81%, negatif 15%, netral 30% sehingga membuktikan sentiment masyarakat pada *platform twitter* mengenai kendaraan listrik tergolong positif.

Kata kunci: *twitter*, *klasifikasi*, *python*, *confusion matrix*.

Abstract: The presence of Electric Vehicles has generated pros and cons among the society, as extensively discussed on the social media platform *Twitter*. Some members of the society argue that Electric Vehicles are claimed to be an effective solution to the fossil fuel deficit, while others believe that a mature preparation, particularly in terms of infrastructure, is necessary to transition to Electric Vehicles. Considering the abundance of opinions on *Twitter*, it is important to perform sentiment classification to facilitate the identification of the prevailing sentiment towards Electric Vehicles, whether they tend to be positive, negative, or neutral. In this study, the researcher utilized the *Twitter platform*, which involved data crawling, cleaning, preprocessing (including case folding, tokenizing, stopword removal), and labeling. *Naive Bayes* was selected as the algorithm due to its highest probability or likelihood values in sentiment classification. The data were tested using a confusion matrix in the Python programming language, using *Google Colaboratory*. The research resulted in sentiment analysis regarding Electric Vehicles with a dataset of 1,301 instances, including 940 positive data, 120 negative data, and 241 neutral data. Based on the testing using

the confusion matrix in Python with Google Colaboratory, an accuracy result of 75% was obtained, with a positive precision value of 81%, negative precision of 15%, neutral precision of 30%, and a positive recall success rate of 75%. These values indicate that the Naive Bayes is considered quite good in processing Twitter documents, with an accuracy percentage of 81% for positive sentiment, 15% for negative sentiment, and 30% for neutral sentiment, thus proving that the sentiment of the Twitter community regarding Electric Vehicles is generally positive.

Keywords: twitter, classification, python, confusion matrix.

1. Pendahuluan

Seiring dengan kemajuan teknologi, khususnya dalam bidang kendaraan seperti kendaraan listrik yang menggantikan yang sebelumnya menggunakan bahan bakar minyak dengan tenaga listrik, perkembangan kendaraan bertenaga listrik saat ini dianggap sebagai langkah awal yang positif dalam upaya mengatasi pencemaran udara global. Meskipun penggunaan kendaraan listrik masih terbatas saat ini, diharapkan bahwa seiring berjalannya waktu, penggunaan kendaraan listrik akan semakin meluas di seluruh dunia, berkontribusi pada peningkatan kualitas udara. [1]

Salah satu layanan *internet* yang paling populer di Indonesia adalah media sosial, dan salah satunya adalah *Twitter*. Di *Twitter*, terdapat istilah "*tweet*" yang memungkinkan pengguna untuk menyampaikan berita terbaru dan pendapat mereka melalui tulisan. Pengguna *Twitter* seringkali membicarakan topik atau hal-hal yang sedang menjadi pembahasan perbincangan utama.[2]

Dengan perkembangan teknologi informasi yang terus meningkat, dampaknya juga dirasakan dalam kemajuan teknologi kendaraan, khususnya kendaraan listrik. Kendaraan yang dulunya menggunakan bahan bakar minyak kini beralih menjadi menggunakan tenaga listrik. Kemajuan kendaraan bertenaga listrik pada saat ini dianggap sebagai langkah awal yang positif untuk mengurangi polusi udara global. Meskipun penggunaan kendaraan listrik masih terbatas saat ini, namun diharapkan bahwa penggunaannya akan semakin meningkat seiring berjalannya waktu. Di Indonesia, sebagai negara yang mendorong akselerasi penggunaan kendaraan listrik, pemasaran kendaraan listrik pada tahun 2021 mencapai 3.193 unit. Namun, terjadi peningkatan yang signifikan sebesar 383,46% pada tahun 2022, dengan sekitar 15.437 unit kendaraan listrik yang berhasil dipasarkan.[3]

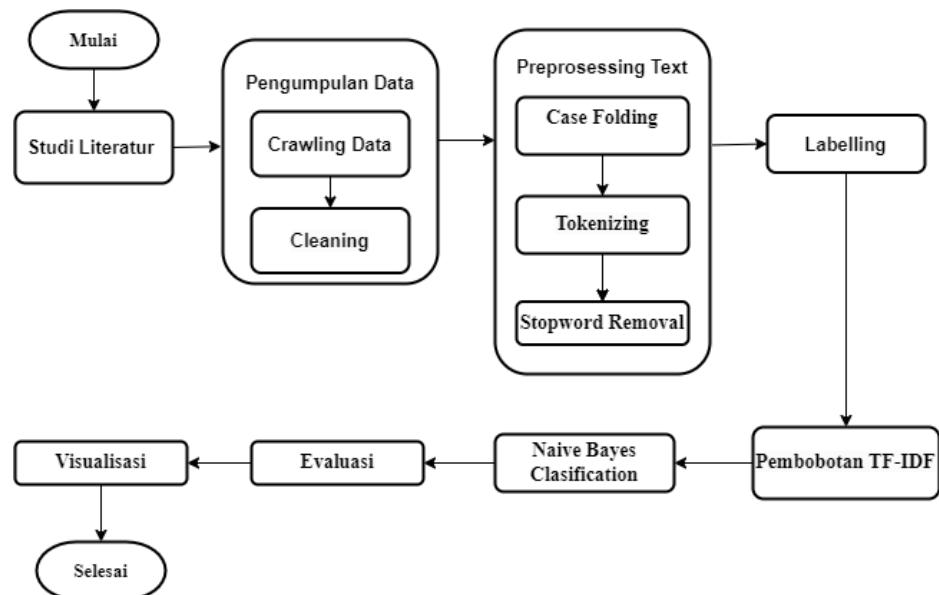
Pemerintah telah mengeluarkan Peraturan Pemerintah (PP) Nomor 55 Tahun 2019 yang mengatur tentang akselerasi program kendaraan listrik berbasis baterai untuk angkutan jalan. Langkah ini memicu perbincangan dan tanggapan dari masyarakat, terutama di media sosial seperti *Twitter*, di mana banyak pendapat dan opini tentang hadirnya kendaraan listrik diungkapkan.[4]

Kehadiran kendaraan listrik menimbulkan pro dan kontra tanggapan masyarakat seperti yang ramai di bicarakan di media sosial *twitter*. Sebagian kalangan masyarakat berpendapat kendaraan listrik di klaim efektif menjadi solusi *deficit* migas tetapi sebagiannya lagi berpendapat untuk menuju kendaraan listrik perlu persiapan yang sangat matang, khususnya pada segi infrastruktur. Banyaknya respon dari masyarakat Indonesia dalam menilai kendaraan listrik menjadi salah satu tolak ukur peneliti untuk melakukan analisis sentimen masyarakat terhadap kendaraan listrik pada media sosial *twitter*. Salah satu metode untuk mengambil kesimpulan tentang pandangan masyarakat Indonesia terhadap kendaraan listrik adalah melalui analisis sentimen. Analisis sentimen, juga dikenal sebagai opini mining, adalah cabang dari klasifikasi teks yang berhubungan dengan pemrosesan bahasa alami, komputasi linguistik, dan penambangan teks secara umum. Tujuan dari analisis sentimen adalah menganalisis pendapat, sentimen, pembicara atau penulis terkait terhadap topik tertentu, serta aktivitas lainnya.[5] Analisis sentimen bermanfaat untuk menentukan apakah opini atau *review* tersebut mempunyai kecenderungan positif, negatif, netral. Metode *Naive Bayes* adalah metode yang cocok untuk digunakan dalam data yang memiliki jumlah yang banyak serta dapat menangani data yang kosong (*missing value*). Selain itu, metode ini juga dapat mengatasi atribut yang tidak seragam dan gangguan yang terdapat dalam data.[6]

Berdasarkan uraian diatas, yang menjadi dasar dalam penelitian ini yaitu bertujuan untuk mengetahui bagaimana sentimen atau opini masyarakat Indonesia terhadap kendaraan listrik, data sentimen yang digunakan berasal dari sosial media *twitter*, kemudian data tersebut nanti akan diklasifikasikan menjadi positif, negatif, dan netral menggunakan metode *Naive Bayes*.

2. Metode Penelitian

Kerangka Penelitian yang telah dibuat dari analisis sentimen terhadap kendaraan listrik menggunakan algoritma Naïve Bayes ini ditunjukkan pada Gambar 1.



Sumber : Hasil Penelitian (2023)

Gambar 1. Kerangka Penelitian.

Kegiatan pada gambar 1. Menunjukkan tahapan penelitian ini yang terdiri dari :

Studi Literatur

Pada tahap awal penelitian hal pertama yang dilakukan yaitu melakukan *review* pada jurnal dari *internet* yang berkaitan dengan topik penelitian terdahulu tentang analisis sentimen. Setelah mereview penelitian terdahulu, dapat dijadikan referensi untuk tahap – tahap penyelesaian yang digunakan untuk melakukan penelitian.

Pengumpulan Data

Crawling Data

Data yang dipergunakan pada penelitian diperoleh melalui pencarian serta pengumpulan data. asal data yang dikumpulkan merupakan *tweet* yang ada di *platform* media sosial *Twitter*. Proses pengumpulan data berasal *Twitter* melakukannya dengan memakai *Twitter* api (*Application Programming Interface*). setelah *tweet* berhasil diambil, data tersebut akan disimpan dalam format *Excel*.

Crawling Data di *Twitter* merujuk pada proses pengambilan atau pengunduhan data dari server *Twitter* dengan menggunakan *Application Programming Interface (API) Twitter*. Data yang dapat diambil meliputi data pengguna (*user*) maupun data *tweet*. Pada proses *crawling data* ini dilakukan dengan menggunakan *keyword* Kendaraan Listrik. *Crawling Data* dilakukan pada tanggal 30 maret 2023 dengan penarikan data dari tanggal 10 maret 2023 dengan menggunakan *Tools RapidMiner* karena pada bulan tersebut munculnya berita pemerintah Indonesia serta Amerika serikat baru saja menandatangani kolaborasi pada banyak sekali sektor energi baru terbaruka, selain untuk mencapai sasaran Indonesia bebas emisi karbon di 2060 kolaborasi juga menyangkut sektor kendaraan listrik.[7] yang menjadikan berita tersebut *trending* topik di *twitter* maupun media *social* lainnya.

Cleaning

Cleaning merupakan tahapan penghapusan data duplikat, dan tahap pembersihan data bertujuan untuk menghapus karakter (simbol), nama *username*, *retweet* (RT). Proses *cleaning* dengan menggunakan *Tools RapidMiner*. [8]

Preprocessing Text

Pada tahap ini, peneliti akan menyiapkan data yang akan diolah. Berikut tahapan *prerocessing* yang akan dilakukan oleh peneliti, dalam penelitian sentimen data *tweet* ini.

Case folding, pada tahapan ini, peneliti akan mengubah seluruh huruf kapital dalam text pada *tweet* sebagai huruf kecil, untuk memudahkan dalam mengidentifikasi kasus dalam proses penyamaan sebuah dokumen. [9]

Tokenizing, pada tahap in, peneliti akan memisahkan atau memotong kalimat *text* menjadi kata perkata, untuk mendapatkan *token*, untuk mengidentifikasi istilah-istilah atau kata-kata yang sering muncul dan tidak *relevan* dengan analisis sentimen ini akan dihilangkan atau dihapus. [10]

Stopword Removal, tahap Stopword Removal melibatkan langkah-langkah untuk menghilangkan kata-kata yang tidak relevan. Kata-kata tersebut tidak memberikan kontribusi signifikan terhadap informasi dalam kalimat. Contohnya adalah kata hubung "yang", "akan", "di", "pada", dan "dan". Dengan menghapus kata-kata ini, informasi dalam kalimat tetap dapat dipertahankan tanpa pengaruh yang signifikan. [11]

Labelling

Dalam melakukan analisis sentimen pada saat ini, metode yang akan digunakan adalah metode berbasis *lexicon* yang menggunakan *library vader sentiment*. Metode berbasis *lexicon* berarti bahwa kata-kata penting dalam dokumen dipilih sesuai dengan kamus/lexicon yang sudah ada. *Vader Sentiment Lexicon* adalah salah satu kamus *lexicon* yang digunakan, kamus ini mengandung 7.500 token yang terdiri dari kata-kata berbahasa Inggris. [12]

Pada tahap saat ini peneliti melakukan instal google translate menggunakan bahasa *python*, yang akan digunakan untuk mentranslate data pada bahasa indonesia ke bahasa inggris, karena untuk melabeli data secara otomatis data harus diubah ke dalam bahasa inggris terlebih dahulu. Ada 3 label yang digunakan dalam penelitian ini yaitu, *positive*, *neutral*, dan *negative*. Untuk menginstall google translate.

Pembobotan Kata (TF-IDF)

Setelah selesai melakukan *preprocessing* (*transformasi*, *tokenisasi*, dan *filtering*) di dokumen, langkah berikutnya adalah melakukan proses term weighting. pada tahap ini, *term* akan diberi bobot atau nilai yang menunjukkan tingkat kepentingannya terhadap dokumen. Dilanjutkan dengan menghitung bobot kepada term yang dicari pada suatu dokumen, hal ini bermaksud agar memahami kegunaan serta persamaan term pada dokumen. Semakin sering sebuah kata yang akan muncul pada setiap dokumen, meningkat pula nilai atau bobot kata tersebut. Sesudah tahap pembobotan terselesaikan, langkah berikutnya adalah proses klasifikasi. Metode yang digunakan buat pembobotan kata adalah metode Tf-Idf. [13]

Naïve Bayes

Dokumen yang relevan dengan kategori tertentu dapat mengalami proses klasifikasi dengan menggunakan kata-kata yang terdapat dalam dokumen tersebut. Mengumpulkan kata-kata dalam dokumen tersebut bertujuan untuk mengkategorikan kata-kata tersebut memiliki nilai makna tertentu yang berkaitan. Salah satu metode yang digunakan untuk mengevaluasi sentimen dari setiap dokumen tweet adalah metode Naive Bayes. [14]

Evaluasi

Evaluasi pada penelitian ini menggunakan *confusion matrix* (untuk pengukuran performa untuk problem klasifikasi *machine learning* dimana keluaran bisa berupa 2 kelas atau lebih). penilaian dilakukan dengan tujuan untuk mengukur kinerja sistem yang telah dibuat berdasarkan hasil klasifikasi yang diperoleh. [15]

Visualisasi

pada tahap visualisasi dengan menggunakan wordcloud untuk setiap sentimen positif, negatif, dan netral berdasarkan data yang telah diproses. Tujuan dari visualisasi ini adalah untuk mendapatkan informasi mengenai kata-kata yang paling sering muncul dalam cuitan.[16]

3. Hasil dan Pembahasan

Pengumpulan Data

Ada 2 tahapan pada proses pengumpulan data yang akan di lakukan oleh peneliti yaitu, proses *crawling* data yang di lakukan pada 30 maret 2023 dengan penarikan data dari 10 maret 2023 menggunakan *tools* rapidminer hasil *crawling* data yang telah dilakukan oleh peneliti. *crawling* data kendaraan listrik menarik data sebanyak 4.261 dataset *twitter* dari limit 5000. Dilanjutkan dengan proses cleaning hasil *Cleaning* data yang telah dilakukan oleh peneliti. *Cleaning* data dengan kendaraan listrik menyisakan data sebanyak 1.301 dataset *twitter* dari limit 4.261.

Preprocessing

Pada tahap ini dilakukan pengolahan terhadap hasil *crawling* yang formatnya tidak terstruktur sehingga menjadi kata dasar. *Preprocessing* dilakukan memakai bantuan *library* di bahasa pemrograman *Python* menggunakan *Google Colaboratory*. Hal pertama yang dilakukan adalah melakukan *case folding*, *tokenizing*, serta *stopword removal* sehingga dapat menghasilkan data bersih dan siap untuk lanjut pada proses berikutnya.

1. Case folding

Proses *case folding* dilakukan bertujuan agar mengubah semua kata dalam teks menjadi huruf kecil. Setelah dilakukan *case folding* data disajikan pada gambar 2.

	Text	CaseFolding
0	Mendukung kendaraan listrik sebagai transporta...	mendukung kendaraan listrik sebagai transporta...
1	supaya beli kendaraan listrik	supaya beli kendaraan listrik
2	kekayaan alam dapat dimanfaatkan dengan dalam ...	kekayaan alam dapat dimanfaatkan dengan dalam ...
3	Semakin maraknya kendaraan listrik ini akan se...	semakin maraknya kendaraan listrik ini akan se...

Sumber : Hasil Penelitian (2023)

Gambar 2 Hasil *Case Folding*

2. Tokenizing

Tokenizing adalah proses untuk memisahkan semua kata dalam setiap dokumen. Setelah dilakukan *tokenizing* data disajikan pada gambar 3.

	Text	CaseFolding	Token
0	Mendukung kendaraan listrik sebagai transporta...	mendukung kendaraan listrik sebagai transporta...	[mendukung, kendaraan, listrik, sebagai, trans...
1	supaya beli kendaraan listrik	supaya beli kendaraan listrik	[supaya, beli, kendaraan, listrik]
2	kekayaan alam dapat dimanfaatkan dengan dalam ...	kekayaan alam dapat dimanfaatkan dengan dalam ...	[kekayaan, alam, dapat, dimanfaatkan, dengan, ...]
3	Semakin maraknya kendaraan listrik ini akan se...	semakin maraknya kendaraan listrik ini akan se...	[semakin, maraknya, kendaraan, listrik, ini, a...]

Sumber : Hasil Penelitian (2023)

Gambar 3 Hasil *Tokenizing*

3. Stopword removal

Stopword Removal merupakan proses untuk menghapus kata-kata yang sering muncul dan kata-kata yang tidak relevan dengan analisis sentimen ini. Hasil *stopword removal* data disajikan pada gambar 4.

	Text	CaseFolding	Token	Stopword
0	Mendukung kendaraan listrik sebagai transporta...	mendukung kendaraan listrik sebagai transporta...	[mendukung, kendaraan, listrik, sebagai, trans...	[mendukung, kendaraan, listrik, transportasi, ...
1	supaya beli kendaraan listrik	supaya beli kendaraan listrik	[supaya, beli, kendaraan, listrik]	[beli, kendaraan, listrik]
2	kekayaan alam dapat dimanfaatkan dengan dalam ...	kekayaan alam dapat dimanfaatkan dengan dalam ...	[kekayaan, alam, dapat, dimanfaatkan, dengan, ...	[kekayaan, alam, dimanfaatkan, menciptakan, in...
3	Semakin maraknya kendaraan listrik ini akan se...	semakin maraknya kendaraan listrik ini akan se...	[semakin, maraknya, kendaraan, listrik, ini, a...	[semakin, maraknya, kendaraan, listrik, semaki...

Sumber : Hasil Penelitian (2023)

Gambar 4 Hasil *Stopword Removal*

Labelling

Proses labelling sentimen pada setiap *tweet* dalam penelitian ini di bagi menjadi positif, negatif, dan netral. Pada melakukan analisis sentimen, metode yang dipergunakan ialah *lexicon based* dengan memakai *library vader sentiment*. *Vader sentiment lexicon* adalah salah satu kamus *lexicon* yang memiliki 7.500 tokens yang mengandung istilah bahasa inggris. Berikut hasil pelebelan ada pada gambar 5.

	Tweet	Compound_Score	Sentiments
0	support environmentally friendly transport ele...	0.7096	Positif
1	buy an electric vehicle	0.0000	Netral
2	Natural resources are utilized to create the e...	0.5574	Positif
3	the more widespread electric vehicles the bett...	0.4404	Positif
4	how dismal is public transportation to be repa...	-0.6124	Negatif
5	electric vehicles destroy nature Indonesia's p...	-0.5423	Negatif

Sumber : Hasil Penelitian (2023)

Gambar 5 Hasil *Labelling*

Pembobotan Kata (TF-IDF)

Tahap selanjutnya pembobotan kata menggunakan metode TF-IDF dapat dirumuskan sebagai berikut.

$$IDF (Word) = \text{Log} \frac{tf}{df} \dots\dots\dots (1)$$

Dokumen yang akan digunakan dalam proses TF-IDF disajikan pada tabel 1.

Tabel 1. Dokumen Pembobotan Kata TF-IDF

Dokumenn	Tweet
D1	mendukung kendaraan listrik transportasi ramah lingkungan
D2	gimana suram transportasi publik diperbaiki malah diperparah subsidi kendaraan listrik
D3	beli kendaraan listrik

Sumber : Hasil Penelitian (2023)

Pembobotan kata dimulai dengan menentukan *Term Frequency* (TF) yang merupakan menghitung jumlah kemunculan kata di setiap dokumen. Kata yang tidak muncul akan bernilai 0, dan jika ada maka bernilai 1. Hasil TF disajikan pada tabel 2.

Tabel 2. Hasil *Term Frequency*

TERM	TF		
	d1	d2	d3
mendukung	1	0	0
kendaraan	1	1	1
listrik	1	1	1
transportasi	1	1	0
ramah	1	0	0
lingkungan	1	0	0
gimana	0	1	0
suram	0	1	0
publik	0	1	0
diperbaiki	0	1	0
malah	0	1	0
diperparah	0	1	0
subsidi	0	1	0
beli	0	0	1

Sumber : Hasil Penelitian (2023)

Setelah nilai TF didapatkan proses selanjutnya adalah mencari nilai *Inverse Document Frequency (IDF)*, dimana IDF merupakan hasil inverse dari DF. Hasil IDF disajikan pada tabel 2.

Tabel 3. Hasil *Inverse Dokumen Frequency*

TERM	DF	IDF
mendukung	1	0,477121
kendaraan	3	0
listrik	3	0
transportasi	2	0,176091
ramah	1	0,477121
lingkungan	1	0,477121
gimana	1	0,477121
suram	1	0,477121
publik	1	0,477121
diperbaiki	1	0,477121
malah	1	0,477121
diperparah	1	0,477121
subsidi	1	0,477121
beli	1	0,477121

Sumber : Hasil Penelitian (2023)

Setelah mendapatkan nilai IDF, langkah berikutnya adalah menghitung nilai TF-IDF. Nilai TF-IDF didapatkan dengan mengalikan *frekuensi* kemunculan kata di setiap dokumen (TF) dengan hasil bobot kata dalam seluruh dokumen (IDF). Hasil dari perkalian antara TF dan IDF disajikan dalam Tabel 4.

Tabel 4. Hasil TF-IDF

TERM	TF-IDF		
	d1	d2	d3
Mendukung	0,477121	0	0
Kendaraan	0	0	0
Listrik	0	0	0
Transportasi	0,176091	0,176091	0
Ramah	0,477121	0	0
Lingkungan	0,477121	0	0
Gimana	0	0,477121	0
Suram	0	0,477121	0
Publik	0	0,477121	0
Diperbaiki	0	0,477121	0
Malah	0	0,477121	0
diperparah	0	0,477121	0
subsidi	0	0,477121	0
beli	0	0	0,477121
Jumlah tf-idf	1,607455	3,51594	0,477121

Sumber : Hasil Penelitian (2023)

Klasifikasi Naïve Bayes

Setelah melewati tahap preprocessing, dokumen tweet dapat diklasifikasikan ke dalam kategori yang sesuai berdasarkan istilah-istilah yang terdapat dalam suatu dokumen tersebut. Kumpulan kata-kata dalam dokumen akan digunakan untuk menentukan kategori yang dianggap memiliki makna yang relevan. Metode klasifikasi Naïve Bayes akan bertujuan untuk digunakan membuat keputusan dengan memprediksi kasus berdasarkan hasil dari klasifikasi yang telah diperoleh.

1) Menghitung Klasifikasi

Berikut ini contoh dari 3 data tweet yang akan digunakan dalam analisis. Setiap data tweet memiliki sampel yang ditandai menjadi 1 positif, 1 negatif, serta 1 netral. Data ini akan dipergunakan sebagai data latih. ditinjau pada tabel 5.

Tabel 5. Data Sampel

Dokumen	Sentimen
mendukung kendaraan listrik transportasi ramah lingkungan	Positif
gimana suram transportasi publik diperbaiki malah diperparah subsidi kendaraan listrik	Negatif
beli kendaraan listrik	Netral

Sumber : Hasil Penelitian (2023)

Pada tabel 5 sampel dataset. Perhitungan dilakukan menggunakan Microsoft Excel, dengan tahap pertama yaitu mengitung *term frequency* (TF) kemunculan kata dalam dokumen yang dapat dilihat pada Tabel 6.

Tabel 6. Term Frequency

TERM	TF		
	d1	d2	d3
mendukung	1	0	0
kendaraan	1	1	1
listrik	1	1	1
transportasi	1	1	0
ramah	1	0	0
lingkungan	1	0	0
gimana	0	1	0
suram	0	1	0
publik	0	1	0
diperbaiki	0	1	0
malah	0	1	0
diperparah	0	1	0
subsidi	0	1	0
beli	0	0	1

Sumber : Hasil Penelitian (2023)

Kemudian, dilakukan perhitungan untuk menentukan jumlah kemunculan kata unik serta jumlah total kata di setiap kategori. Alur perhitungan tersebut bisa dilihat dalam Tabel 7 pada bawah ini.

Tabel 7. Kemunculan Kata Unik

Kategori	Jumlah dokumen di setiap kategori	jumlah dokumen seluruhnya	P(Kategori)	Jumlah Kata Setiap Kategori
Positif	1	3	0,3333333	6
Negatif	1	3	0,3333333	10
Netral	1	3	0,3333333	3
Jumlah kata unik		14		

Sumber : Hasil Penelitian (2023)

Kata unik merujuk pada kata-kata yang muncul secara konsisten pada setiap dokumen. Selanjutnya selesai mendapatkan hasil jumlah kata unik dan total kata dalam setiap kategori, langkah selanjutnya ialah menghitung probabilitas untuk setiap kategori.

$$P(\text{Kata}) = \frac{\text{jumlah dari satu kata di satu kategori} + 1}{\text{jumlah kata tiap kategori} + \text{jumlah keseluruhan kata unik}} \dots\dots\dots (2)$$

Pada tabel *term frequency* (TF) di atas, diperoleh hasil dari jumlah kata dalam masing-masing kategori sebagai berikut: 6 kata dalam kategori positif, 10 kata dalam kategori negatif, dan 3 kata dalam kategori netral. Selain itu, terdapat 14 kata unik yang muncul di keseluruhan dataset.

2) Pengujian Naïve Bayes menggunakan bahasa pemrograman *Python*

Model yang akan digunakan untuk melakukan klasifikasi merupakan alur pengujian *Naive Bayes*. Proses ini dilakukan dengan memanfaatkan *library* pada bahasa pemrograman *Python*. Untuk proses klasifikasi dapat dilihat pada gambar 6.

```
#Training Model

from datetime import datetime
start_time = datetime.now()
from joblib import dump
#algoritme fitting

text_algorithm = GaussianNB ()

model = text_algorithm.fit(X_train, y_train)

# save the model to disk
dump(model, filename="model_sentiment_naive.joblib")

end_time = datetime.now()
result_time =end_time-start_time
print("Duration:",result_time)
```

Sumber : Hasil Penelitian (2023)

Gambar 5. *Import Library*

Proses klasifikasi melibatkan perhitungan probabilitas untuk setiap kalimat terhadap setiap kelas, sehingga menghasilkan hasil yang jelas mengenai data yang dimasukkan. Terdapat 261 data uji yang akan digunakan untuk menguji model klasifikasi yang telah dibuat menggunakan metode Naïve Bayes.

Evaluasi

Evaluasi bertujuan untuk mengevaluasi model klasifikasi dan memeriksa kebenaran atau kesalahan prediksinya. Untuk mengukur performa Algoritma Naïve Bayes, akan dilakukan pengujian pada model yang telah dilakukan. Hasil pada klasifikasi dievaluasi menggunakan *confusion matrix*. *Confusion matrix* digunakan sebagai metode untuk mengukur sejauh mana kinerja suatu metode klasifikasi dan memberikan informasi tentang *akurasi*, *presisi*, dan *recall*. Tabel 8 menampilkan hasil *confusion matrix* setelah pengujian terhadap seluruh data *tweet* menggunakan bahasa pemrograman *Python* dengan menggunakan *Google Colaboratory*.

Tabel 8. *Confusion Matrix*

		Prediksi		
		Negatif	Netral	Positif
Aktual	Negatif	3	4	24
	Netral	1	12	16
	Positif	16	24	182

Sumber : Hasil Penelitian (2023)

Penghitungan Akurasi

$$\text{Accuracy} = \frac{TP + TN + TI}{\text{Jumlah Data}} \times 100\% \dots\dots\dots (3)$$

$$= \frac{182 + 3 + 12}{261} = \frac{197}{261} = 0,75 = 75\%$$

Penghitungan Precision Positif

$$Precision\ Positif = \frac{TP}{TP + FP1 + FP2} \times 100\% \dots\dots\dots (4)$$

$$= \frac{182}{182 + 24 + 16} = \frac{182}{222} = 0,81 = 81\%$$

Penghitungan Precision Negatif

$$Precision\ Negatif = \frac{TN}{TN + FN1 + FN2} \times 100\% \dots\dots\dots (5)$$

$$= \frac{3}{3 + 16 + 1} = \frac{3}{20} = 0,15 = 15\%$$

Penghitungan Precision Netral

$$Precision\ Netral = \frac{TI}{TI + FI1 + FI2} \times 100\% \dots\dots\dots (6)$$

$$= \frac{12}{12 + 24 + 4} = \frac{12}{40} = 0,3 = 30\%$$

Perhitungan manual Recall Positif

$$Recall\ Positif = \frac{TP}{FN1 + FI1 + TP} \times 100\% \dots\dots\dots (7)$$

$$= \frac{182}{16 + 24 + 182} = \frac{182}{222} = 0,81 = 81\%$$

Perhitungan manual Recall Negatif

$$Recall\ Negatif = \frac{TN}{FP1 + FI2 + TN} \times 100\% \dots\dots\dots (8)$$

$$= \frac{3}{24 + 4 + 3} = \frac{3}{31} = 0,09 = 9\%$$

Perhitungan manual Recall Netral

$$Recall\ Netral = \frac{TI}{FN2 + FP2 + TI} \times 100\% \dots\dots\dots (9)$$

$$= \frac{12}{1 + 16 + 12} = \frac{12}{29} = 0,41 = 41\%$$

1) Visualisasi

Setelah proses pembersihan data serta klasifikasi dilakukan, diperoleh hasil sentimen. *Wordcloud* digunakan sebagai bentuk visualisasi data teks yang menampilkan deretan kata yang paling sering muncul pada analisis teks tersebut. *Script Python* untuk membuat *wordcloud* bisa ditinjau di Gambar 7



Gambar 6 Word cloud

Sumber : hasil penelitian (2023)

4. Kesimpulan

Berdasarkan hasil penelitian mengenai analisis sentimen kendaraan listrik pada platform Twitter, terdapat sebanyak 1.301 data yang telah melewati tahap preprocessing teks, yang terdiri dari *case folding*, *tokenizing*, *stopword removal*, serta *klasifikasi* menggunakan algoritma naïve bayes dan evaluasi data menggunakan confusion matrix bahasa pemrograman *Python* memakai *Google Colaboratory*, serta disimpulkan bahwa hasil menunjukkan bahwa tanggapan masyarakat mengenai kendaraan listrik menghasilkan sebanyak 1.301 data yang telah dilabelling dengan masing-masing data sentimen positif 940 data, sentimen Negatif 120 data, sentimen Netral 241 data. Dengan hasil persentase 75% pada akurasi yaitu artinya seberapa besar tingkat akurasi model yang telah dibuat dapat mengklasifikasi data yang benar. Hasil menunjukkan bahwa *Precision* menggambarkan keakuratan data yang diminta dengan hasil diberikan oleh model, positif sebesar 81%, negatif sebesar 15%, dan netral sebesar 30%, dan tingkat keberhasilan *Recall* menggambarkan sejauh mana keberhasilan model pada menemukan kembali informasi yang dimasukkan pada pengujian Positif sebesar 81%, negatif sebesar 15%, dan netral sebesar 30% Berdasarkan nilai tersebut membuktikan sentiment masyarakat mengenai kendaraan listrik pada platform twitter menghasilkan sentimen positif.

Referensi

- [1] Y. Pratama, D. Triantoro Murdiansyah, and K. Muslim Lhaksana, "Analisis Sentimen Kendaraan Listrik Pada Media Sosial Twitter Menggunakan Algoritma Logistic Regression dan Principal Component Analysis," *Jurnal Media Informatika Budidarama*, pp. 529–535, 2023, doi: 10.30865/mib.v7i1.5575.
- [2] D. Darwis, E. Shintya Pratiwi, and A. Ferico Octaviansyah Pasaribu, "PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA," *Jurnal Ilmiah Edutic*, vol. 7, no. 1, 2020.
- [3] R. Mustajab, "Penjualan Mobil Listrik di Indonesia," *DataIndonesia.id*. Accessed: May 11, 2023. [Online]. Available: <https://dataindonesia.id/sektor-riil/detail/penjualan-mobil-listrik-di-indonesia-capai-15437-unit-pada-2022>
- [4] A. F. Riyadi, F. R. Rahman, M. A. Nofa Pratama, M. K. Khafidli, and H. Patria, "Pengukuran Sentimen Sosial Terhadap Teknologi Kendaraan Listrik: Bukti Empiris di Indonesia," *EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi*, vol. 11, no. 2, p. 141, Dec. 2021, doi: 10.36448/expert.v11i2.2171.
- [5] F. Mailo, "Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia," *Jurnal Sistem Informasi Kesehatan Masyarakat Journal of Information Systems for Public Health*, vol. 4, no. 1, 2019.
- [6] T. Arifin, "PREDIKSI PENYAKIT GINJAL KRONIS MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER BERBASIS PARTICLE SWARM OPTIMIZATION," *Jurnal Tekno Insentif*, vol. 13, no. 1, pp. 26–30, Apr. 2019, doi: 10.36787/jti.v13i1.97.
- [7] F. Fansuri, "RI Gaet AS Percepat EBT Hingga Kendaraan Listrik," *CNBC Indonesia TV*, *CNBC Indonesia*. Accessed: Mar. 26, 2023. [Online]. Available: <https://www.cnbcindonesia.com/news/20230317172534-8-422702/ri-gaet-as-percepat-ebt-hingga-kendaraan-listrik>
- [8] W. Yulita *et al.*, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *JDMSI*, vol. 2, no. 2, pp. 1–9, 2021.
- [9] E. Febriyani, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Algoritma Naive Bayes Classifier Di Twitter," *Jurnal TEKNO KOMPAK*, vol. 17, no. 1, 2019.
- [10] R. B. Kurniawan, A. Suwarisman, I. Afriyanti, A. Wahyudi, and D. D. Saputra, "Analisis Sentimen Complain dan Bukan Complain pada Twitter Telkomsel dengan SMOTEdan Naïve Bayes," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 7, no. 1, 2023, doi: 10.35870/jti.
- [11] P. Pandunata, S. R. Ali, and Y. Nurdiansyah, "ANALISIS SENTIMEN PROGRAM MERDEKA BELAJAR KAMPUS MERDEKA MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER SENTIMENT ANALYSIS OF MERDEKA BELAJAR KAMPUS MERDEKA USING NAÏVE BAYES CLASSIFIER ALGORITHM," *Jurnal Sistem Informasi Dan Bisnis Cerdas*, vol. 16, no. 1, 2023.

- [12] F. Rejeki, "Analisa Sentimen Mengenai Kenaikan Harga Bbm Menggunakan Metode Naïve Bayes Dan Support Vector Machine," *JSAI : Journal Scientific and Applied Informatics*, vol. 6, no. 1, 2023, doi: 10.36085.
- [13] W. T. Nonia, R. Indriati, and M. Najibullo Muzak, "Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter," *Jambura Journal of Electrical and Electronics Engineering*, vol. 3, no. 2, 2021, [Online]. Available: <https://t.co/TIk5mK5bwS>
- [14] M. J. Dwi, "Analisis Perbandingan Algoritma Klasifikasi untuk Penyakit Jantung," *Journal of Informatics and Computer Science*, vol. 04, 2023.
- [15] O. I. Gifari, M. Adha, I. Rifky Hendrawan, and F. Freddy Setlight Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY)*, vol. 2, no. 1, 2022.
- [16] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.