

# PERBANDINGAN METODE NAÏVE BAYES DAN SUPPORT VECTOR MACHINE (SVM) PADA ANALISIS SENTIMEN KENDARAAN LISTRIK PADA MEDIA SOSIAL “X”

Dani Saputra<sup>1</sup>, Agus Wantoro<sup>2</sup>, Damayanti<sup>3</sup>, Permata<sup>4</sup>, Rusliyawati<sup>5</sup>

<sup>1,2,3,4,5</sup> Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia

<sup>1</sup>dani\_saputra@teknokrat.ac.id, <sup>2</sup>aguswantoro@teknokrat.ac.id, <sup>3</sup>damayanti@teknokrat.ac.id,

<sup>4</sup>permata@teknokrat.ac.id, <sup>5</sup>rusliyawati@teknokrat.ac.id

## Abstrak

Kendaraan listrik merupakan kendaraan yang menggunakan satu atau lebih motor listrik atau motor traksi sebagai tenaga penggerak. Kendaraan listrik saat ini menjadi tren global sebagai alternatif kendaraan berbahan bakar fosil atau bahan bakar minyak (BBM). Peralihan dari kendaraan berbahan bakar minyak ke kendaraan listrik menjadi salah satu solusi efektif karena kendaraan listrik memiliki beberapa keunggulan dibanding kendaraan BBM. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kemunculan kendaraan listrik pada media “X” dan melakukan perbandingan kinerja algoritma klasifikasi. Data sentimen selanjutnya akan dilakukan klasifikasi menggunakan metode *Naïve Bayes* dan *Support Vector Machine* (SVM). Pengelolaan data menggunakan bahasa pemrograman *python* dan *google Collaboratory*. Tahapan penelitian ini meliputi *crawling*, *labelling*, *pre-processing*, klasifikasi metode, dan visualisasi. Berdasarkan 2615 data yang telah dikumpulkan, selanjutnya dilakukan *preprocessing*, dan *cleaning data* sehingga. Hasil *cleaning* didapatkan 1245 data yang akan dilakukan klasifikasi. Dataset dibagi menjadi dua, yaitu data latih dan data uji. Berdasarkan data latih, selanjutnya dilakukan pelabelan data menghasilkan 57,16% data positif sebanyak 30,02 data netral, dan 17,83 data negatif. Hasil perbandingan klasifikasi menunjukkan bahwa metode *Naïve Bayes* memiliki akurasi sebesar 66%, sedangkan metode *Support Vector Machine* (SVM) sebesar 93%. Metode *Support Vector Machine* (SVM) terbukti lebih unggul untuk analisis data dibandingkan dengan metode *Naïve Bayes*. Selain itu berdasarkan data pengguna “X” lebih banyak yang memberikan respon positif terhadap kendaraan listrik

**Kata kunci** : Kendaraan Listrik, Naïve Bayes, Support Vector Machine, Analisis Sentimen, “X”

## 1 Pendahuluan

Kendaraan listrik telah menjadi topik yang semakin penting dalam dunia otomotif selama beberapa tahun terakhir. Kendaraan listrik merupakan jenis kendaraan yang menggunakan listrik sebagai sumber tenaga utama untuk menggerakkan roda (Alfarizi & Fitriani, 2023). Kendaraan listrik merupakan salah satu sarana transportasi yang bisa mengurangi pemakaian energi fosil agar memenuhi kebutuhan mobilitas masyarakat namun tetap ramah lingkungan karena tidak memiliki polusi (Aryanti & Santoso, 2023). Sehingga peralihan kendaraan bisa menjadi solusi yang efektif karena kendaraan listrik memiliki banyak keunggulan, seperti lebih ramah lingkungan, efisiensi energi, dan biaya operasional yang lebih rendah (Ernawati et al., n.d.). Kendaraan listrik menjadi salah satu alternatif yang menjanjikan dalam mengatasi masalah lingkungan dan semakin meningkatnya harga. Kelebihan kendaraan listrik terletak pada minimnya komponen dibandingkan dengan kendaraan konvensional, sehingga perawatannya lebih mudah. Diharapkan bahwa dengan keunggulan-keunggulannya, kendaraan listrik dapat secara signifikan menggantikan kendaraan konvensional yang bergantung pada BBM (Kusuma et al., 2023). Kendaraan listrik merupakan produk

ekonomi sirkulasi dan juga teknologi baru di dunia otomotif. Produk ekonomi sirkulasi semakin populer sebagai model alternatif yang mengurangi penipisan sumber daya, pemborosan, dan emisi di bidang akademik (Ananda et al., 2023). Meningkatnya penggunaan kendaraan berbahan bakar minyak yang berdampak pada pencemaran udara dan iklim global mendorong pemerintah membuat kebijakan untuk beralih menjadi kendaraan berbasis energi listrik. Meski kendaraan listrik sudah menjadi isu dunia dalam menekan pemanasan global tetapi belum menjadi alternatif pilihan utama (Alfarizi et al., 2024).

Pertama kali mobil listrik diproduksi secara massal dan komersial oleh *General Motors* yang dirancang untuk digunakan pada durasi waktu yang terbatas dan menempuh jarak yang relatif pendek daripada dengan kendaraan konvensional. (Caswadi et al., 2024). Namun masih terdapat permasalahan seperti infrastruktur yang belum memadai, harga yang relatif mahal, dan waktu pengisian baterai yang cukup lama.

Saat ini masyarakat belum mengetahui apakah munculnya kendaraan listrik memberikan sentimen negatif atau positif. Selain itu, klasifikasi sentimen dapat dilakukan menggunakan beberapa metode klasifikasi untuk memberikan informasi berupa metode yang lebih baik antara *Naïve Bayes* dan

*Support Vector Machine*, terhadap dalam analisis sentimen kendaraan listrik.

Metode *naïve bayes* dikenal dengan metode yang relatif sederhana namun efektif, terutama untuk klasifikasi teks seperti analisis sentiment, dan *naïve bayes* juga dapat bekerja dengan baik ketika data latih relatif lebih kecil, karena tidak memerlukan sejumlah data yang besar untuk menghasilkan prediksi yang kuat. Sedangkan metode *support vector machine* dikenal dengan menghasilkan akurasi yang tinggi, terutama dalam tugas klasifikasi yaitu analisis sentimen, dan *support vector machine* juga dapat bekerja sangat baik dengan data yang kompleks.

Menurut penelitian yang dilakukan oleh Putri Gea Aryanti dan Imam Santoso dengan judul “Analisis Sentimen pada “X” Terhadap Mobil Listrik Menggunakan Algoritma *Naïve Bayes*” peneliti mengambil data tweets sebanyak 700, pengujian hanya menggunakan Algoritma *Naïve Bayes* hasilnya terdapat 650 data tweets yang mengandung sentimen positif dan 50 data tweets mengandung sentimen negatif. Dalam penelitiannya mendapatkan nilai akurasi sebesar 87,43% dan AUC 0,518% (Aryanti & Santoso, 2023). Menurut penelitian yang dilakukan oleh Ni Wayan Ernawati, I Nyoman Satya Kumara, dan Widyadi Setiawan dengan judul “Perbandingan Metode Klasifikasi *Support Vector Machine* Dan *Naïve Bayes* Pada Analisis Sentimen Kendaraan Listrik” peneliti mengambil data tweets sebanyak 3.582, berdasarkan hasil Metode *Naïve Bayes* memperoleh hasil 82% dan SVM memperoleh hasil 81% (Ernawati et al., n.d.). Menurut penelitian yang dilakukan oleh Salman Alfarizi dan Eka Fitriani dengan judul “Analisis Sentimen Kendaraan Listrik Menggunakan Algoritma *Naïve Bayes* dengan Seleksi *Fitur Information Gain* dan *Particle Swarm Optimization*) peneliti mengambil data tweets sebanyak 600. berdasarkan hasil menggunakan algoritma *Naïve Bayes* mendapatkan akurasi 79,43 dan nilai AUC 0,639 (Alfarizi & Fitriani, 2023). Menurut penelitian yang dilakukan oleh Adittia Agustian, Tukino, dan Fitria Nurapriani dengan judul “Penerapan Analisis Sentimen dan *Naïve Bayes* Terhadap Opini Penggunaan Kendaraan Listrik Di “X”” peneliti mengambil data tweets sebanyak 1589, berdasarkan hasil pengujian algoritma *Naïve Bayes* menghasilkan positif sebesar 42,98%, netral sebesar 52,02%, dan negatif sebesar 0,5% (Agustian et al., 2022). Menurut penelitian yang dilakukan oleh Ike Verawati dan Syarif Nurwahid Jaelani dengan judul “Analisis Sentimen Pengguna “X” Terhadap Bus Listrik Menggunakan *Naïve Bayes*” peneliti mengambil data tweets sebanyak 2585, berdasarkan hasil pengujian diperoleh kelas sentiment terbanyak adalah sentimen positif sebanyak 77,32%, sentimen negatif sebanyak 22,69% (Verawati & Jaelani, 2024).

Pengambilan data dari media sosial “X” memiliki informasi yang lebih relevan untuk tujuan penelitian yang dilakukan dan datanya lebih mudah diakses atau lebih memiliki kualitas yang lebih baik

dibandingkan dengan media sosial lainnya seperti Facebook, YouTube, dan sebagainya.

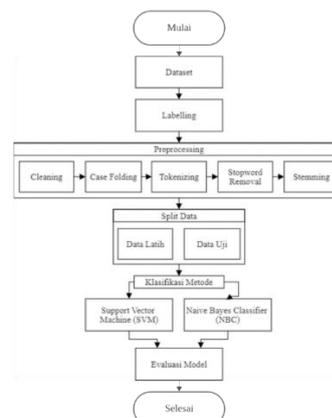
Kebaruan penelitian ini bahwa belum ditemukan oleh peneliti sebelumnya tentang analisis sentimen kendaraan listrik pada media “X” menggunakan menggunakan algoritma *Naïve Bayes*, dan *Support Vector Machine* (SVM). Analisis sentimen ini dibagi menjadi tiga kelas yaitu sentimen positif, netral, dan negatif. Adapun data yang digunakan berasal dari “X” dengan kata kunci kendaraan listrik. Penelitian ini akan memberikan pengetahuan kepada produsen untuk melakukan evaluasi berdasarkan persepsi masyarakat terhadap penggunaan kendaraan listrik

## 2 Metode

Tahapan kerangka penelitian dari analisis sentimen terhadap kendaraan listrik menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*.

Tahapan penelitian dimulai dengan melakukan pengumpulan dataset yang diambil dari “X” dengan menggunakan Teknik *crawling* dengan menggunakan bahasa pemrograman *python* memanfaatkan token API. Setelah mendapatkan data tersebut, selanjutnya dilakukan tahap *labelling* secara manual membagi kategori sentimen, yaitu sentimen positif dan negatif. Selanjutnya dilakukan tahapan *preprocessing*, proses ini mengubah data yang tidak terstruktur menjadi terstruktur sebagai bahan penelitian. Tahapan *preprocessing* yaitu *cleaning*, *case folding*, *tokenizing*, *stop word removal* dan *stemming*.

Selanjutnya tahap *split* data, dimana tahapan ini dibagi menjadi data *training* dan data *testing*. Data *training* adalah digunakan untuk melatih algoritma untuk mengidentifikasi data yang termasuk dalam sentimen positif dan negatif, sedangkan data *testing* digunakan untuk menguji model yang diperoleh dari data training. Tahap selanjutnya adalah klasifikasi metode yaitu dibagi menjadi *Naïve Bayes* dan *Support Vector Machine*. Tahap yang terakhir yaitu evaluasi model, evaluasi model dilakukan dengan menggunakan perhitungan *confusion matrix* untuk mendapatkan nilai performa *accuracy*, *precision*, *recall* dan *f1-score*. Untuk kerangka penelitian dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Penelitian

Kerangka penelitian pada Gambar 1 digunakan sebagai dasar tahapan penelitian yang akan dilakukan.

## 2.1 Pengumpulan Data (Collecting)

Proses pengumpulan data diambil dari media sosial “X” dengan menggunakan *Application Programming Interface* (API). Beberapa tahapan crawling data ditampilkan pada Gambar 2.



Gambar 2. Tahapan Crawling Data

Pengambil data dari media sosial X dikarenakan memiliki informasi yang lebih relevan dengan tujuan dari penelitian yang dilakukan dan lebih mudah diakses atau lebih memiliki kualitas yang lebih baik daripada media sosial lain. Data yang dapat diambil berupa data user tweet. Crawling data tweet dilakukan dengan menggunakan keyword “Kendaraan Listrik”. Pengumpulan data menggunakan *tools Google Collaboratory* dengan memanfaatkan token API “X”, yang dimana token tersebut dapat diperoleh pada setiap akun “X” penggunaanya (Alin et al., 2023).

## 2.2 Labelling

*Labelling* atau pelabelan data adalah proses pemberian label atau kategori data mentah untuk mendukung pelatihan *machine learning*, dengan pembagian kategori sentimen yaitu sentimen positif, netral dan negatif (Farhani & Sutisna, 2024).

## 2.3 Pre-processing

*Pre-processing* adalah serangkaian langkah yang dilakukan untuk membersihkan, mengatur, dan mempersiapkan data teks mentah agar dapat diolah lebih lanjut oleh pemrosesan bahasa alami atau model pembelajaran mesin lainnya. Langkah-langkah dalam *preprocessing* bervariasi tergantung pada kebutuhan dan tujuan spesifik dari analisis (Tarigan & Yusupa, 2024). Beberapa langkah umum dalam *preprocessing* sebagai berikut:

1. *Cleaning*, yaitu data yang dilakukan pembersihan dengan tujuan untuk menghapus tanda baca, *mention*, *hashtag*, *link*, dan karakter lainnya.
2. *Case Folding*, proses dalam preprocessing teks dimana semua karakter huruf diubah menjadi huruf kecil atau huruf besar.
3. *Tokenizing*, proses ini yaitu memisahkan teks menjadi unit-unit yang lebih kecil seperti kata-kata, frasa, atau kalimat.
4. *Stepword Removal*, proses ini berfungsi untuk menghilangkan kata-kata yang tidak penting, seperti “yang”, “di”, dan lain-lain. Tahapan ini sangat berpengaruh terhadap klasifikasi metode yang akan digunakan.

5. *Stemming*, proses ini berfungsi untuk mengubah kata yang berimbuhan menjadi kata dasar.

## 2.4 Split Data

*Dataset* dibagi menjadi dua yaitu data *training* dan data *testing*. Data *training* digunakan untuk melatih algoritma untuk mengidentifikasi data yang termasuk dalam sentimen positif atau negatif. Setelah selesai melakukan data *training*, data *testing* digunakan untuk menguji model yang diperoleh dari data *training*. Persentase data *training* 70% dan data *testing* 30%.

## 2.5 Klasifikasi

Klasifikasi adalah pengelompokan berupa *predictive* data mining. Dalam *predictive* data mining, terdapat proses yang disebut data *splitting*. Data *splitting* tahapan membagi menjadi dua yaitu data *training* dan data *testing*. Data *training* untuk melatih sistem, sedangkan data *testing* untuk menguji sistem. Algoritma NBC dan SVM yang digunakan dalam penelitian ini termasuk dalam *supervised learning* yang berarti data digunakan untuk melatih model yang sudah diberi label. Dalam analisis ini, label yang digunakan adalah sentimen positif dan negatif.

## 2.6 Naïve Bayes Classifier (NBC)

Algoritma NBC menggunakan probabilitas dan statistik untuk menyelesaikan masalah klasifikasi. Metode ini mengklasifikasikan dengan menghitung nilai probabilitas  $P(X|Y)$  (Putro et al., 2020). Metode NBC banyak digunakan karena relatif sederhana namun efektif, terutama untuk klasifikasi teks seperti analisis sentimen. Metode NBC dapat bekerja dengan baik ketika data latih relatif lebih kecil, karena tidak memerlukan sejumlah data yang besar untuk menghasilkan prediksi yang kuat. Keuntungan dari klasifikasi ini adalah kebutuhan akan data pelatihan yang relatif sedikit untuk memperkirakan parameter yang diperlukan. Persamaan (1) untuk menghitung nilai  $P(X|Y)$ .

$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)} \quad (1)$$

Pada persamaan 1 mulai dari fungsi  $P(X|Y)$  merupakan *posterior* | *probability* yaitu nilai X berdasarkan kondisi Y. Variabel  $P(X|Y)$  merupakan probabilitas Y yang ditentukan X adalah benar, sedangkan  $P(X)$  adalah peluang evidence dari nilai X, dan  $P(Y)$  probabilitas dari nilai Y.

## 2.7 Support Vector Machine (SVM)

SVM merupakan metode klasifikasi yang menggunakan *supervised learning* untuk memprediksi kelas berdasarkan pola dari proses pelatihan. Klasifikasi dilakukan dengan *hyperplane* yang memisahkan kelas positif dan kelas negatif.

*Hyperplane* yang optimal memiliki jarak terjauh ke titik data pelatihan terdekat dari setiap kelas, karena margin yang lebih besar biasanya mengurangi *error* generalisasi. Pemisahan terbaik dilakukan dengan mengukur margin *hyperplane* dan menemukan titik maksimalnya (Apriyani & Kurniati, 2020). SVM dikenal akurasi yang tinggi, terutama dalam tugas klasifikasi yaitu analisis sentimen. SVM dapat bekerja sangat baik dengan data yang kompleks. Persamaan (2) perhitungan dari metode SVM untuk menentukan suatu klasifikasi atau prediksi.

$$f(X_d) = \sum_{i=1}^{ns} a_i y_i x_i x_d + b \quad (2)$$

Persamaan (2) dimana variabel *ns*= jumlah *support vector*, *a<sub>i</sub>*= nilai bobot setiap titik data, *y<sub>i</sub>*= kelas data, *x<sub>i</sub>*= *variable support vector*, *x<sub>d</sub>*= data yang akan diklasifikasikan, dan *b*= nilai *error*

### 2.8. Evaluasi Model

Tahap evaluasi bertujuan untuk menilai kinerja model algoritma yang diterapkan. Evaluasi model dilakukan dengan menggunakan tabel *confusion matrix*, yaitu sebuah tabel yang menyediakan informasi perbandingan antara hasil klasifikasi yang dihasilkan oleh sistem berupa prediksi, dan hasil klasifikasi yang sebenarnya (Fikri et al., 2020). Tabel *confusion matrix* menunjukkan jumlah data uji yang diklasifikasi dengan benar dan jumlah data uji yang diklasifikasikan dengan salah (Maulana et al., 2024).

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Presisi = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP+TN}{TP+FN} \quad (5)$$

$$f1 - score = \frac{2 \times Presisi \times Recall}{Presisi \times Recall} \quad (6)$$

Variabel *true positif* (TP) dan *true negatif* (TN), menunjukkan klasifikasi dilakukan dengan benar. *False positif* (FP) adalah prediksi yang menunjukkan hasil positif tapi sebenarnya negatif, sedangkan *false negatif* (FN) adalah situasi dimana prediksi menunjukkan hasil negatif tapi seharusnya positif. Tabel 1 menunjukkan matriks dari confusion matrix

**Tabel 1.** Confusion Matrix

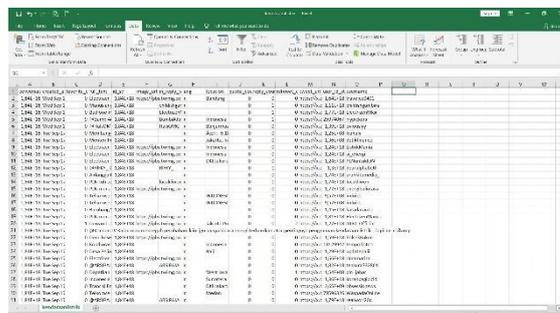
Kelas Prediksi	Kelas Aktual	
	Sentimen Positif	Sentimen Negatif
Sentimen Positif	TP	FN
Sentimen Negatif	FP	TN

Berdasarkan confusion matrix pada Tabel 1, maka tabel ini akan digunakan untuk mengukur kinerja algoritma Support Vector Machine (SVM) dan Naive Bayes Classification (NBC) untuk mencari sentimen positif dan sentimen negatif dari kelas prediksi dan kelas aktual.

## 3 Hasil dan Pembahasan

### 3.1 Pengumpulan Dataset

Pengambilan data untuk analisis sentimen diperoleh dengan cara *crawling* dengan menggunakan bahasa pemrograman *python* melalui token API “X”. Gambar 3 menampilkan hasil *crawling dataset*.

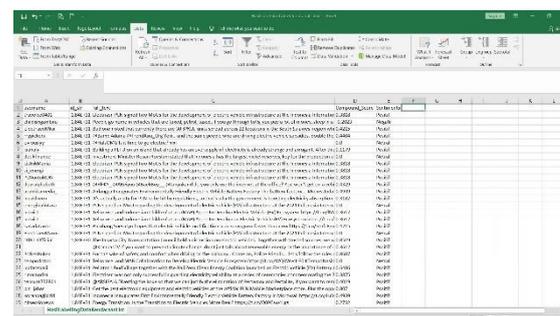


Gambar 3. Hasil Crawling Data

Gambar 3 menampilkan bahwa setiap satu kode token dapat mengumpulkan data sebanyak 1.500 setiap harinya. Dataset yang telah di *crawling* menggunakan token API terkumpul sebanyak 2.615, data tersebut masih berupa data mentah yang harus melalui tahap *filtering* dan *preprocessing* terlebih dahulu agar data dapat dipakai untuk analisis sentimen secara lebih akurat

### 3.2 Labelling Dataset

Proses ini dilakukan untuk menentukan kelas sentimen yang akan digunakan untuk menghitung nilai akurasi dan *visualisasi* data tweet. Sebelum proses pelabelan dimulai, data harus diterjemahkan ke dalam bahasa Inggris menggunakan analisis sentimen *Vader* agar hasilnya optimal. Hasil *labelling* dapat dilihat pada Gambar 4.



Gambar 4. Hasil Labelling

Hasil pelabelan dibagi menjadi tiga kategori yaitu sentimen positif, negatif, dan netral, namun sentimen netral tidak digunakan karena dianggap tidak bermakna atau tidak memiliki pengaruh signifikan

### 3.3 Preprocessing Dataset

#### 1. Cleaning

Tahap *cleaning* dilakukan untuk membersihkan dokumen dari isi data yang tidak

penting atau tidak berguna untuk tahap pengujian berikutnya. Berikut ini merupakan hasil dari tahap *cleaning* pada Tabel 2.

**Tabel 2.** Hasil Cleaning

Tweet	Cleaning
Electrizen PLN signed two MoUs for the development of electric vehicle infrastructure at the Indonesia International Sustainability Forum 2024. This MoU shows PLN's commitment to accelerating the EV ecosystem in Indonesia. #PLN #PLNForIndonesia #AcceleratingRenewableEnergy <a href="https://t.co/07zyIuCmYN">https://t.co/07zyIuCmYN</a>	Electrizen PLN signed two MoUs for the development of electric vehicle infrastructure at the Indonesia International Sustainability Forum This MoU shows PLN's commitment to accelerating the EV ecosystem in Indonesia PLN #PLNForIndonesia AcceleratingRenewableEnergy

Berdasarkan pada tabel *cleaning*, maka tabel ini bertujuan untuk membersihkan data yang berantakan seperti hastag dan url.

2. Case Folding

Tahapan ini berfokus dalam transformasi kalimat yang mana huruf besar dirubah menjadi huruf kecil. Sehingga memastikan konsistensi format huruf, Hasil proses *case folding* dapat dilihat pada Tabel 3.

**Tabel 3.** Hasil Case Folding

Tweet	Case Folding
Electrizen PLN signed two MoUs for the development of electric vehicle infrastructure at the Indonesia International Sustainability Forum 2024. This MoU shows PLN's commitment to accelerating the EV ecosystem in Indonesia. #PLN #PLNForIndonesia #AcceleratingRenewableEnergy <a href="https://t.co/07zyIuCmYN">https://t.co/07zyIuCmYN</a>	electrizen pln signed two mous for the development of electric vehicle infrastructure at the indonesia international sustainability forum this mou shows plns commitment to accelerating the ev ecosystem in indonesia pln plnforindonesia acceleratingrenewableenergy

Berdasarkan dari tabel case folding, hasil dari tabel ini bertujuan untuk mentransformasi kalimat seperti huruf besar dirubah menjadi huruf kecil semua.

3. Tokenizing

*Tokenizing* adalah proses dimana kalimat dipecah menjadi beberapa kata yang terpisah. Tahapan ini bertujuan untuk membagi teks menjadi kata-kata individu dengan menghapus karakter pemisah seperti titik, koma, spasi, dan lain sebagainya. Hasil proses *tokenizing* dapat dilihat pada Tabel 4.

**Tabel 4.** Hasil Tokenizing

Tweet	Tokenizing
Electrizen PLN signed two MoUs for the development of electric vehicle infrastructure at the Indonesia International Sustainability Forum 2024. This MoU shows PLN's commitment to accelerating	['electrizen', 'pln', 'signed', 'two', 'mous', 'for', 'the', 'development', 'of', 'electric', 'vehicle', 'infrastructure', 'at', 'the', 'indonesia', 'international', 'sustainability', 'forum', 'this', 'mou', 'shows',

the EV ecosystem in Indonesia. #PLN #PLNForIndonesia #AcceleratingRenewableEnergy <a href="https://t.co/07zyIuCmYN">https://t.co/07zyIuCmYN</a>	['plns', 'commitment', 'to', 'accelerating', 'the', 'ev', 'ecosystem', 'in', 'indonesia', 'pln', 'plnforindonesia', 'acceleratingrenewableenergy']
-------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------

Hasil dari tabel *tokenizing*, bertujuan untuk membagi teks keseluruhan menjadi kata-kata individu dengan menghapus karakter seperti koma, titik, spasi, dan lainnya.

4. Stop Word Removal

Proses ini berfungsi untuk menghapus kata-kata atau kalimat yang tidak penting atau tidak memiliki makna dalam suatu data. Hasil proses *stop word removal* dapat dilihat pada Tabel 5.

**Tabel 5.** Hasil Stop Word Removal

Tweet	Stop Word
Electrizen PLN signed two MoUs for the development of electric vehicle infrastructure at the Indonesia International Sustainability Forum 2024. This MoU shows PLN's commitment to accelerating the EV ecosystem in Indonesia. #PLN #PLNForIndonesia #AcceleratingRenewableEnergy <a href="https://t.co/07zyIuCmYN">https://t.co/07zyIuCmYN</a>	['electrizen', 'pln', 'signed', 'two', 'mous', 'for', 'the', 'development', 'of', 'electric', 'vehicle', 'infrastructure', 'at', 'the', 'indonesia', 'international', 'sustainability', 'forum', 'this', 'mou', 'shows', 'plns', 'commitment', 'to', 'accelerating', 'the', 'ev', 'ecosystem', 'in', 'indonesia', 'pln', 'plnforindonesia', 'acceleratingrenewableenergy']

Berdasarkan hasil dari tabel *stop word removal*, bertujuan untuk menghapus kata atau kalimat yang tidak memiliki makna dalam suatu data seperti yang, dan, di, ke.

5. Stemming

Tahap ini berguna untuk menghilangkan imbuhan pada kata atau kata dasar. Hasil proses *stemming* dapat dilihat pada Tabel 6.

**Tabel 6.** Hasil Stemming

Tweet	Stemming
Electrizen PLN signed two MoUs for the development of electric vehicle infrastructure at the Indonesia International Sustainability Forum 2024. This MoU shows PLN's commitment to accelerating the EV ecosystem in Indonesia. #PLN #PLNForIndonesia #AcceleratingRenewableEnergy <a href="https://t.co/07zyIuCmYN">https://t.co/07zyIuCmYN</a>	electrizen pln signed two mous for the development of electric vehicle infrastructure at the indonesia international sustainability forum this mou shows plns commitment to accelerating the ev ecosystem in indonesia pln plnforindonesia acceleratingrenewableenergy

Berdasarkan dari tabel *stemming*, hasil dari tabel ini bertujuan untuk meningkatkan pemrosesan teks dalam machine learning agar kata-kata yang berbeda akan menjadi satu kata dasar.

3.4. Evaluasi dan Klasifikasi Metode

Setelah melakukan tahap *preprocessing* data selanjutnya penelitian ini mengeluarkan hasil

implementasi dengan metode *naïve bayes*. Dengan jumlah data yang di uji yang menghasilkan nilai akurasi, *precision*, *recall*, *support*, dan *f1-score* yang terlihat pada Gambar 5.

Laporan Klasifikasi :				
	precision	recall	f1-score	support
Negatif	0.17	0.26	0.20	89
Positif	0.83	0.74	0.78	439
accuracy			0.66	528
macro avg	0.50	0.50	0.49	528
weighted avg	0.72	0.66	0.69	528

Gambar 5. Hasil Klasifikasi Naive Bayes

Berdasarkan Gambar 5, dapat diketahui nilai akurasi yang diperoleh dari data kendaraan listrik sebesar 66% dengan *precision* negatif 17%, *precision* positif 83%, *recall* negatif 26%, *recall* positif 74%, *f1-score* negatif 20%, *f1-score* positif 78% dengan *support* 528. Klasifikasi dilakukan dengan data latih sebanyak 1246 dan data uji sebanyak 535. Hasil perhitungan dengan metode SVM dengan topik kendaraan listrik dan data yang serupa dengan fungsi *confusion matrix* dapat dilihat pada Gambar 6.

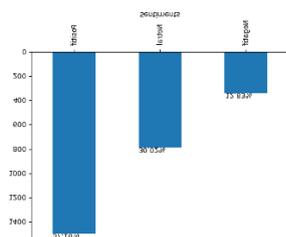
	precision	recall	f1-score	support
Negatif	0.95	0.61	0.74	95
Positif	0.92	0.99	0.96	440
accuracy			0.93	535
macro avg	0.94	0.80	0.85	535
weighted avg	0.93	0.93	0.92	535

Gambar 6. Hasil Klasifikasi Support Vector Machine

Berdasarkan hasil perhitungan Gambar 6, nilai akurasi pada topik kendaraan listrik dengan metode SVM memiliki nilai yang lebih besar dibandingkan menggunakan metode NBC dengan jumlah dan data yang sama. Nilai akurasi SVM sebesar 93% dengan *precision* negatif 95%, *precision* positif 92%, *recall* negatif 61%, *recall* positif 99%, *f1-score* negatif 74%, *f1-score* positif 96% dengan *support* 535

### 3.5. Visualisasi

Pada tahap *visualisasi* akan menampilkan hasil dari *histogram*, *wordcloud*, *visualiasi* sentimen kata negatif, dan positif. Analisis sentimen menggunakan pemrograman *pyhton* dengan menampilkan hasil *histogram* sentimen positif, negatif, dan netral dari *dataset* yang sudah melalui proses pelabelan menggunakan *vader* sentimen. Visualisasi histogram ditampilkan pada Gambar 7.

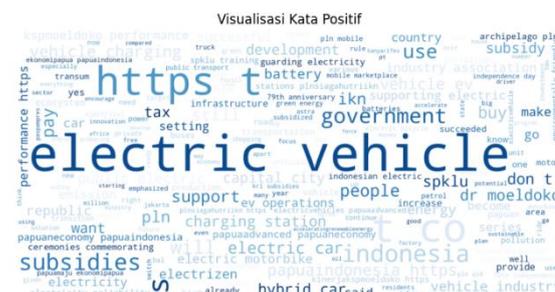


Gambar 7. Histogram Sentimen

Berdasarkan Gambar 7 berupa sentimen histogram dapat dilihat bahwa sebesar 57,16% dengan kategori Positif, dan 30,02% kategori Netral, dan 12,83% untuk kategori Negatif. Kategori Positif berdasarkan data sentiment dikarenakan kendaraan listrik sangat menghemat energi dan dapat mengurangi emisi karbon. Kategori Netral menurut data yaitu cenderung bertanya terkait apakah saatnya beralih ke kendaraan listrik dan untuk kategori Negatif karena dampak buruknya pengembangan kendaraan listrik dapat memicu krisis air dan sarana dan prasarannya kurang memadai.

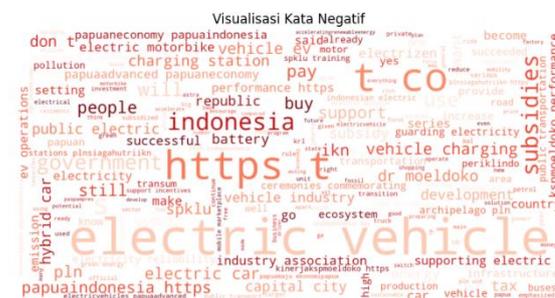
### 3.6. Word Cloud

Tahapan *word cloud* menampilkan hasil berupa kata-kata yang dianalisis sentimen berupa bentuk pengeluaran kata negatif dan positif. Hasil *word cloud* ini sama dengan *histogram* menggunakan bahasa *python* dengan bantuan *tools google Collaboratory* sebagai pencariannya. Hasil *visualisasi* kata-kata positif yang paling sering muncul seperti *electric vehicle*, *tax*, *government*, *battery*, dan sebagainya. Dapat dilihat hasil *word cloud visualisasi* kata positif pada Gambar 8.



Gambar 8. Word Cloud Visualisasi Kata Positif

Hasil *visualisasi* menunjukkan kata-kata negatif yang paling sering muncul seperti *vehicle charging*, *subsidies*, *electric vehicle*, *republic*, dan sebagainya. Untuk hasil *visualisasi* kata negatif dapat dilihat pada Gambar 9.



Gambar 9. Word Cloud Visualisasi Kata Negatif

Selain menampilkan *word cloud visualisasi* kata positif dan negatif, proses ini menghasilkan *visualisasi* dari *wordcloud* keseluruhan data yang mana setiap kata pada data tersebut akan tervisualisasikan. Hasil proses *word cloud* dilihat pada Gambar 10.



- KLASIFIKASI SUPPORT VECTOR MACHINE DAN NAÏVE BAYES PADA ANALISIS SENTIMEN KENDARAAN LISTRIK. *Jurnal SPEKTRUM*, Vol. 10, N.
- Farhani, A., & Sutisna. (2024). Analisis Sentimen Terhadap Kendaraan Listrik di Indonesia Menggunakan Metode Klasifikasi Naïve Bayes. *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, 5(3), 2680–2690. <https://doi.org/10.35870/jimik.v5i3.983>
- Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen “X”. *SMATIKA JURNAL*, 10(02), 71–76. <https://doi.org/10.32664/smatika.v10i02.455>
- Kusuma, G. H., Permana, I., Salisah, F. N., Afdal, M., Jazman, M., & Marsal, A. (2023). Pendekatan Machine Learning: Analisis Sentimen Masyarakat Terhadap Kendaraan Listrik Pada Sosial Media X. *JUSIFO (Jurnal Sistem Informasi)*, 9(2), 65–76. <https://doi.org/10.19109/jusifo.v9i2.21354>
- Maulana, B. A., Fahmi, M. J., Imran, A. M., & Hidayati, N. (2024). Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 375–384. <https://doi.org/10.57152/malcom.v4i2.1206>
- Putro, H. F., Vlandari, R. T., & Saptomo, W. L. Y. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi Dan Komunikasi (TIKOMSiN)*, 8(2). <https://doi.org/10.30646/tikomsin.v8i2.500>
- Tarigan, V. T., & Yusupa, A. (2024). Perbandingan Algoritma Maching Learning dalam Analisis Sentimen Mobil Listrik di Indonesia pada Media Sosial “X”/X. *Jurnal Informatika Polinema*, 10(4), 479–490. <https://doi.org/10.33795/jip.v10i4.5130>
- Verawati, I., & Jaelani, S. N. (2024). Analisis Sentimen Pengguna “X” Terhadap Bus Listrik Menggunakan Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(2), 832. <https://doi.org/10.30865/mib.v8i2.7030>