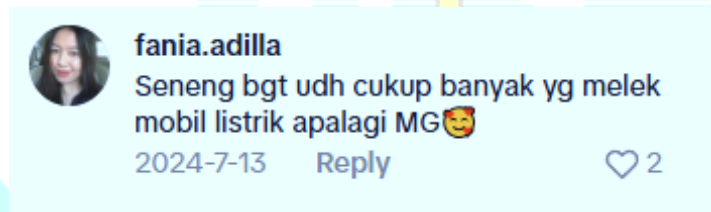


BAB III

METODE PENELITIAN

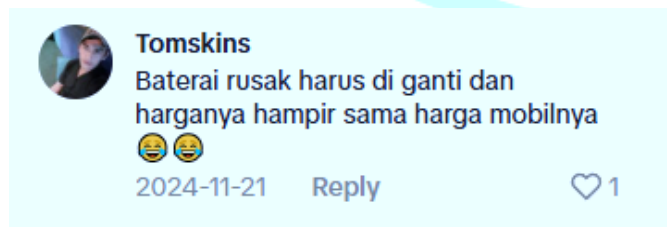
3.1. Objek Penelitian

Objek dalam penelitian ini adalah data komentar publik yang diperoleh dari platform TikTok melalui pencarian dengan menggunakan kata kunci “Mobil Listrik.” Teknik pengumpulan data yang digunakan adalah metode *crawling*, dengan total data yang berhasil diperoleh sebanyak 2.147 komentar. Proses pengumpulan data dilaksanakan dalam dua tahap, yaitu tahap pertama pada tanggal 31 Maret 2024 dan tahap kedua pada tanggal 11 hingga 15 Desember 2024. Data hasil *crawling* terbagi ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral.



Gambar 3. 1 Komentar Positif

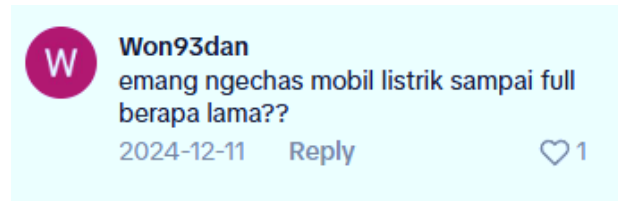
Gambar 3.1 menampilkan contoh komentar yang termasuk ke dalam kategori sentimen positif. Pada komentar tersebut, pengguna mengungkapkan apresiasi terhadap meningkatnya kesadaran masyarakat mengenai keberadaan mobil listrik. Ungkapan seperti "Seneng bgt" dan "udh cukup banyak yg melek mobil listrik" mencerminkan ekspresi positif serta antusiasme terhadap perkembangan mobil listrik.



Gambar 3. 2 Komentar Negatif

Gambar 3.2 menampilkan contoh komentar yang tergolong ke dalam kategori sentimen negatif. Pada komentar tersebut, pengguna menyampaikan keluhan terkait biaya penggantian baterai mobil listrik yang dinilai hampir sebanding dengan harga

mobil itu sendiri. Ungkapan seperti "Baterai rusak harus diganti dan harganya hampir sama harga mobilnya" mencerminkan persepsi negatif terhadap aspek biaya perawatan mobil listrik.

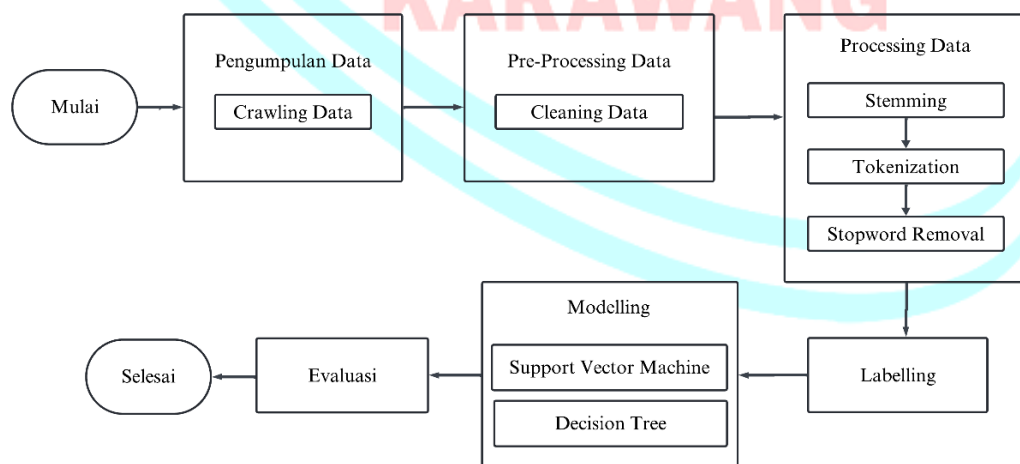


Gambar 3. 3 Komentar Netral

Gambar 3.3 menampilkan contoh komentar yang tergolong ke dalam kategori sentimen netral. Pada komentar tersebut, pengguna mengajukan pertanyaan mengenai durasi waktu yang diperlukan untuk mengisi daya mobil listrik hingga penuh. Ungkapan seperti "emang ngechas mobil listrik sampai full berapa lama?" menunjukkan bahwa komentar tersebut bersifat informatif tanpa mengandung muatan opini positif maupun negatif terhadap mobil listrik.

3.2. Prosedur Penelitian

Penelitian ini meliputi enam tahapan utama, yaitu Pengumpulan Data (*Crawling*), *Pre-Processing Data*, *Preprocessing Data*, *Labelling*, *Modelling*, dan Evaluasi. Adapun prosedur penelitian ini disajikan pada Gambar 3.4.



Gambar 3. 4 Prosedure Penelitian

3.2.1. Pengumpulan Data

Pada tahap ini, pengumpulan data dilakukan menggunakan metode *crawling*. Teknik ini digunakan untuk mengumpulkan komentar dari platform

TikTok berdasarkan kata kunci "Mobil Listrik". Proses *crawling* menghasilkan sebanyak 2.147 komentar yang diperoleh dari dua periode pengumpulan data, yaitu pada tanggal 31 Maret 2024 dan 11–15 Desember 2024. Seluruh data yang diperoleh dari kedua periode tersebut kemudian disimpan dalam format Excel. Adapun hasil dari proses *crawling* data tersebut dapat dilihat pada Gambar 3.5.

dataset-sentiment-mobil-listrik1 (2).xlsx - Excel

File Home Insert Page Layout Formulas Data Review View Developer Help Nitro Pro Tell me what you want to do

Clipboard Paste Font Alignment Number Styles Conditional Formatting Format as Table Cell Styles Insert

Calibri 12 A A Wrap Text B I U Font Merge & Center General % +0.00 -0.00 Conditional Formatting Format as Table Cell Styles Insert

GET GENUINE OFFICE Your license isn't genuine, and you may be a victim of software counterfeiting. Avoid interruption and keep your files safe with genuine Office today. Get genuine

E14 10-3

	A	B	C	D	E
1	e1g2efjf5 href 1zpj2q-lmgAvatar	username	full_text	esugaz-SpanCreated	
2	https://www.tik	https://p16-sign-	apasihh457	Klo di jual lg anjlok bgt hrg nya mbl listrik kya ionic	10-3
3	https://www.tik	https://p16-sign-	Whiplash	Jujur trust issue naek mobil listrik takut habis baterai di	10-4
4	https://www.tik	https://p16-sign-	ewt	Kayanya mobil listrik cocok untuk dalkot aja	10-3
5	https://www.tik	https://p16-sign-	moaisonrin	Aku dah pernah pakai mobil listrik Depok-Bromo malang kak, deg-degan tp seruuu 🤔apalagi pas di tengah sawah, pengalaman tak terlupakan 😊	10-3
6	https://www.tik	https://p16-sign-	Vicky	Tergantung sama waktunya aja kayanya klo buru-buru dan jarak jauh kurang rekomend mobil listrik. Tapi kalo	Oct-20
7	https://www.tik	https://p16-sign-	yudis,anto555	diesel 🤔	10-18
8	https://www.tik	https://p16-sign-	Selvy Tst	saya keluar kota mending pke bensin,klw baterai ngeri² sedep apa lagi suka PP JKT Bandung sendri,kena mobil listrik ngeri² sedap Krn pas diJln macet langsung panik batrenya mau habis 🤔 tp memang bagus sih	10-16
9	https://www.tik	https://p16-sign-	Scorpio	kalau mobil listrik, yang nilang polisi atau PLN kak	10-3
10	https://www.tik	https://p16-sign-	SANTO		10-14
11	https://www.tik	https://p16-sign-	Indra Nugroho	Mobil listrik keknya cocok buat dalam kota doang si menurut saya tergantung peruntukannya, klo dalkot aja	10-4

Gambar 3. 5 Dataset Hasil *Crawling* Komentar TikTok

Gambar 3.5 menunjukkan hasil *crawling* komentar TikTok yang digunakan dalam penelitian ini. Dataset tersebut difokuskan pada dua atribut utama, yaitu *username* yang memuat nama pengguna, serta *full_text* yang berisi isi komentar terkait "Mobil Listrik".

3.2.2. Pre-Processing Data

Proses *pre-processing* data dilakukan untuk mempersiapkan data mentah sehingga layak digunakan dalam analisis lebih lanjut (Ryadi *et al.* 2021). Data mentah yang diperoleh pada tahap *crawling* akan melalui proses pembersihan, yang mencakup penghapusan komentar yang tidak relevan, nilai *null*, data duplikat serta *spam*. Selanjutnya, Pembersihan juga meliputi

penghapusan elemen seperti simbol, angka, *emoticon* dan tagar. Selain itu, seluruh teks dalam komentar dikonversi menjadi huruf kecil *lowercase*.

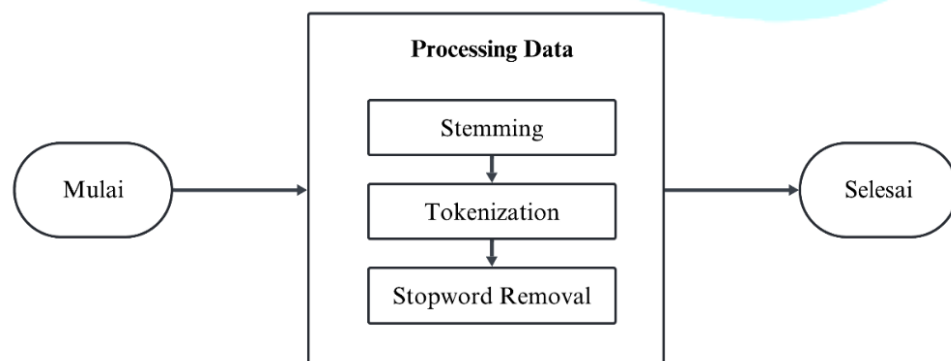
Tabel 3. 1 Contoh Hasil *Pre-Processing*

No	Sebelum	Sesudah
1.	Bensin kali ya kaaa 😭 takut urgent2 nyaaa 😁	bensin kali ya kaaa takut urgent nyaaa
2.	LINGKUNGAN LAH	lingkungan lah
3.	enak BBM mas....gak ruwet	enak bbm mas gak ruwet
4.	Untuk saat ini, lebih masuk akal pilih mobil hybrid.	untuk saat ini lebih masuk akal pilih mobil hybrid
5.	Kenapa POLUSINYA tdk di BAHAS ?	Kenapa polusi nya tdk di bahas

Tabel 3.1 menampilkan contoh hasil tahap *pre-processing* yang dilakukan terhadap data komentar yang telah dikumpulkan. Tujuan dari tahap ini adalah untuk memastikan konsistensi format data serta meningkatkan kualitas data sebelum dilakukan analisis lebih lanjut.

3.2.3. *Processing Data*

Tahap *Processing Data* merupakan langkah pengolahan komentar dalam dataset untuk mengubah data menjadi format yang lebih standar, umum, dan dasar (Kurniawan and Muliawan 2024). Tahap ini bertujuan untuk meningkatkan kualitas data, sehingga data yang dihasilkan layak digunakan dalam proses analisis lebih lanjut. Pada tahap *processing data* terdapat tiga proses utama yang akan dilakukan, sebagaimana ditunjukkan pada Gambar 3.6.



Gambar 3. 6 Tahap *Processing Data*

a. Stemming

Pada Tahap *Stemming*, dilakukan proses normalisasi kata dengan mengubahnya ke bentuk dasar dengan penghapusan kata imbuhan menggunakan *library sastrawi*. Proses ini bertujuan untuk menyatukan variasi kata yang memiliki makna serupa ke dalam satu bentuk dasar, sehingga memudahkan identifikasi dan mengurangi pengaruh variasi bahasa dalam analisis.

Tabel 3. 2 Contoh *Stemming*

Sebelum				Sesudah			
Mobil	listrik	sangat	ramah	Mobil	listrik	sangat	ramah
lingkungan dan efisien digunakan.				lingkungan dan efisien guna.			

b. Tokenization

Tokenization merupakan proses memecah teks atau data menjadi bagian-bagian yang lebih kecil. Pada proses ini, teks dari sebuah kalimat dipecah menjadi potongan-potongan kata. Proses ini bertujuan untuk mempermudah analisis atau pemrosesan lebih lanjut.

Tabel 3. 3 Contoh *Tokenization*

Sebelum				Sesudah			
Mobil	listrik	sangat	ramah	'Mobil',	'listrik',	'sangat',	'ramah',
lingkungan dan efisien guna.				'lingkungan', 'dan', 'efisien', 'guna'			

c. Stopword Removal

Stopword Removal, melibatkan proses menyaring kata-kata untuk mengidentifikasi apakah terdapat kata yang tidak memiliki nilai (*values*) atau tidak relevan, karena tidak memberikan kontribusi informasi pada suatu kalimat. Kata-kata tersebut kemudian dihilangkan dari teks, seperti kata sambung 'dan', 'yang', 'ke', 'dari', dan sebagainya.

Tabel 3. 4 Contoh *Stopword Removal*

Sebelum				Sesudah			
'Mobil',	'listrik',	'sangat',	'ramah',	Mobil',	'listrik',	'sangat',	'ramah',
'lingkungan', 'dan', 'efisien', 'guna'				'lingkungan', 'efisien', 'guna'			

3.2.4. *Labelling*

Labelling atau pemberian label, adalah proses yang dilakukan untuk menentukan respons cuitan komentar dalam dataset secara manual atau otomatis (Santoso *et al.* 2022). Pada tahap ini, setiap cuitan komentar diberi label untuk mengklasifikasikan komentar ke dalam tiga kategori, yaitu positif, negatif, dan netral. Proses pelabelan ini tidak hanya bertujuan untuk mengelompokkan komentar berdasarkan sentimen, tetapi juga untuk membangun dasar representasi fitur teks melalui metode *Term Frequency-Inverse Document Frequency* (TF-IDF).

Setiap kata dalam komentar dihitung bobotnya menggunakan TF-IDF untuk merepresentasikan tingkat kepentingan kata terhadap keseluruhan korpus data. Selain itu, untuk memastikan validitas dan konsistensi label, proses pelabelan divalidasi oleh ahli bahasa, yaitu Delvia Pepy Anggraeni, S.Pd. (Sarjana Sastra Indonesia). Validasi ini bertujuan untuk meminimalisasi potensi kesalahan dalam proses pelabelan serta memastikan keakuratan setiap label yang telah diberikan.

3.2.5. *Modelling*

Pada tahap pemodelan, data komentar yang telah melalui proses pra-pemrosesan dan pelabelan akan digunakan untuk pengujian model. Proses pemodelan dilakukan dengan menerapkan algoritma *Support Vector Machine Linear* dan *Decision Tree*. Data dibagi menjadi dua kelompok, yaitu data latih dan data uji, dengan rasio 80:20. Pemilihan kedua algoritma ini didasarkan pada keunggulannya dalam mengelola dan mengklasifikasikan data yang kompleks, serta didukung penelitian sebelumnya yang membuktikan efektivitasnya dalam analisis sentimen.

3.2.5.1. Penerapan *Support Vector Machine*

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang bekerja dengan mencari garis atau bidang pemisah *hyperplane* terbaik untuk membedakan data ke dalam kelas yang berbeda (Ramadhani *et al.* 2024). Dalam konteks analisis sentimen, SVM digunakan untuk

memisahkan komentar berdasarkan polaritas sentimennya, seperti sentimen positif, negatif, atau netral. Proses kerja SVM dalam klasifikasi analisis sentimen dapat dijelaskan melalui langkah-langkah berikut:

1. Input Data

Data komentar yang telah melalui tahapan pra-pemrosesan digunakan sebagai input. Contoh komentar:

- A. "Mobil listrik ini sangat ramah lingkungan." Sebagai komentar Positif
- B. "Harga mobil listrik masih terlalu mahal." Sebagai komentar Netral
- C. "Baterai mobil cepat habis, tidak efektif." Sebagai komentar Negatif

2. Representasi Fitur (Vektorisasi)

Setiap komentar dikonversi menjadi vektor numerik berdasarkan kata-kata penting. Sebagai contoh.

Tabel 3. 5 Contoh Representasi Fitur

Kata Kunci	Mobil	Ramah	Mahal	Baterai	Efektif
Komentar 1	1	1	0	0	0
Komentar 2	1	0	1	0	0
Komentar 3	1	0	0	1	1

Keterangan:

1 = Kata tersebut ada dalam komentar.

0 = Kata tersebut tidak ada dalam komentar.

3. Pelabelan

Setiap komentar diberi label sentimen sebagai contoh:

- A. 2 = Positif
- B. 1 = Netral
- C. 0 = Negatif

4. Pelatihan Model

Model SVM dilatih dengan data latih tersebut menggunakan *kernel linear*. Tujuan pelatihan adalah mencari *hyperplane* terbaik yang memisahkan komentar berdasarkan label sentimen.

Secara prinsip:

- A. *Hyperplane* adalah garis (di 2D) atau bidang (diruang berdimensi lebih tinggi) yang membagi data berdasarkan kelasnya.
- B. SVM mencari margin maksimal antara kelas-kelas.

5. Klasifikasi Data Uji

Komentar baru akan diproyeksikan ke ruang fitur dan diklasifikasikan berdasarkan sisi *hyperplane* tempat komentar tersebut berada.

6. Hasil Prediksi

Setiap komentar kemudian diklasifikasikan ke dalam kategori sentimen negatif, netral, atau positif. Contoh hasil klasifikasi sentimen menggunakan model SVM disajikan pada Tabel 3.6.

Tabel 3. 6 Contoh Hasil Klasifikasi Sentimen Menggunakan SVM

No	Komentar	Representasi Fitur	Label Asli (Target)	Prediksi SVM
1	Mobil listrik ini sangat ramah lingkungan	[1, 1, 0, 0, 0]	2 (Positif)	2 (Positif)
2	Harga mobil listrik masih terlalu mahal	[1, 0, 1, 0, 0]	1 (Netral)	1 (Netral)
3	Baterai mobil cepat habis, tidak efektif	[1, 0, 0, 1, 1]	0 (Negatif)	0 (Negatif)

3.2.5.2. Penerapan Decision Tree

Decision Tree adalah metode *machine learning* yang digunakan sebagai alat pendukung keputusan, dengan struktur pohon untuk klasifikasi atau regresi (Rokhman *et al.* 2021). Decision Tree dalam analisis sentimen membentuk struktur pohon, dimana setiap *node internal* merepresentasikan pengujian terhadap fitur tertentu, cabang menggambarkan hasil pengujian, dan daun (*leaf node*) menunjukkan label kelas sentimen. Proses kerja Decision Tree dalam klasifikasi analisis sentimen dapat dijelaskan melalui langkah-langkah berikut:

1. Input Data

Data komentar yang telah melalui tahapan pra-pemrosesan digunakan sebagai input. Contoh komentar:

- A. "Mobil listrik ini sangat ramah lingkungan." Sebagai komentar Positif
- B. "Harga mobil listrik masih terlalu mahal." Sebagai komentar Netral
- C. "Baterai mobil cepat habis, tidak efektif." Sebagai komentar Negatif

2. Representasi Fitur (Vektorisasi)

Setiap komentar dikonversi menjadi vektor numerik berdasarkan kata-kata penting. Sebagai contoh.

Tabel 3. 7 Contoh Representasi Fitur

Kata Kunci	Mobil	Ramah	Mahal	Baterai	Efektif
Komentar 1	1	1	0	0	0
Komentar 2	1	0	1	0	0
Komentar 3	1	0	0	1	1

Keterangan:

1 = Kata tersebut ada dalam komentar.

0 = Kata tersebut tidak ada dalam komentar.

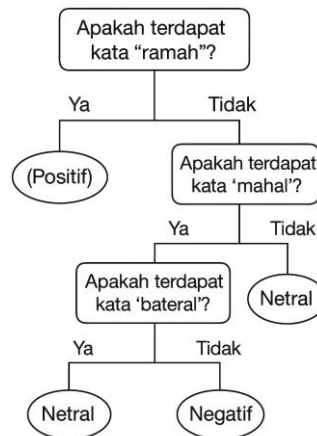
3. Pelabelan

Setiap komentar diberi label sentimen sebagai contoh:

- D. 2 = Positif
- E. 1 = Netral
- F. 0 = Negatif

4. Pelatihan Model Decision Tree

Model Decision Tree dilatih dengan data latih. Algoritma ini secara bertahap memilih fitur yang paling membedakan label sentimen, berdasarkan kriteria tertentu (seperti *Gini Impurity* atau *Entropy*). Contoh cara kerja Decision Tree untuk klasifikasi sentimen ditunjukkan pada Gambar 3.7.



Gambar 3. 7 Ilustrasi Pohon Keputusan Decision Tree

Pada Gambar 3.7 komentar yang mengandung kata "ramah" diklasifikasikan sebagai Positif. Jika tidak terdapat kata "ramah" tetapi terdapat kata "mahal", komentar diklasifikasikan sebagai Netral. Jika hanya ditemukan kata "baterai", komentar diklasifikasikan sebagai Negatif. Jika ketiga kata tersebut tidak ditemukan, komentar diklasifikasikan sebagai Netral.

5. Klasifikasi Data Uji

Komentar baru akan diuji melalui cabang-cabang keputusan dalam pohon. Model akan mengikuti jalur berdasarkan ada/tidak adanya kata kunci tertentu sampai menghasilkan prediksi sentimen.

6. Hasil Prediksi

Setiap komentar diklasifikasikan menjadi salah satu dari tiga label sentimen: negatif, netral, atau positif. Berikut Contoh hasil Decision Tree dalam klasifikasi sentimen yang disajikan pada Tabel 3.7.

Tabel 3. 8 Contoh Hasil Klasifikasi Sentimen Menggunakan Decision Tree

No	Komentar	Representasi Fitur	Label Asli (Target)	Prediksi Decision Tree
1	Mobil listrik ini sangat ramah lingkungan	[1, 1, 0, 0, 0]	2 (Positif)	2 (Positif)
2	Harga mobil listrik masih terlalu mahal	[1, 0, 1, 0, 0]	1 (Netral)	1 (Netral)
3	Baterai mobil cepat habis, tidak efektif	[1, 0, 0, 1, 0]	0 (Negatif)	0 (Negatif)

3.2.6. Evaluasi

Proses evaluasi model dilakukan untuk menilai kinerja dari model yang telah dibangun pada tahap *modelling*. Evaluasi model dilakukan dengan memanfaatkan *confusion matrix* untuk menilai sejauh mana model klasifikasi mampu memprediksi label dengan akurat, yaitu dengan membandingkan hasil prediksi dengan label sebenarnya pada berbagai kategori.

Evaluasi ini bertujuan untuk mengidentifikasi kekuatan dan kelemahan model dalam membedakan antar kategori sentimen, serta menjadi dasar untuk melakukan perbaikan atau penyempurnaan model di tahap selanjutnya. Melalui *confusion matrix*, sejumlah metrik evaluasi, seperti akurasi, *Precision*, *Recall* , dan *F1-score*, dihitung untuk memberikan penilaian komprehensif terhadap performa model dalam mengklasifikasikan sentimen (Safitri and Sasongko 2024).

