

## BAB III

### METODOLOGI PENELITIAN

#### 3.1. Objek Penelitian

Penelitian ini akan berfokus pada analisis sentimen terkait opini atau pendapat pengguna Youtube melalui video review mengenai mobil listrik yang sudah ada di Indonesia. Pada akhir tahun 2024 lalu, terdapat dua mobil listrik dengan jumlah penjualan terbanyak yaitu Byd M6 dengan penjualan mencapai lebih dari 6.000 unit dan pada urutan kedua yaitu Wuling Binguo Ev dengan total penjualan lebih dari 5.000 unit mobil, sedangkan jenis mobil listrik lainnya yang terjual jumlahnya tidak mencapai 200 unit. Oleh karena itu data yang akan digunakan dalam penelitian ini adalah opini masyarakat pada video review mobil Byd M6 dan Wuling Binguo Ev. Berikut merupakan tabel dari detail data yang akan digunakan:

Tabel 3. 1 Dataset yang diambil pada tanggal 22 Februari 2025

Kanal Youtube	Judul Video	Jumlah Komentar	Link
Ridwan Hanif	Wuling Binguo Ev: Kalau Air Ev Kececilan, Beli Ini Aja!	1.549	<a href="https://youtu.be/LdfYG68ya8U?si=Y8DxMo_Wx8ryy8f3">https://youtu.be/LdfYG68ya8U?si=Y8DxMo_Wx8ryy8f3</a>
Fitra Eri	Horee!! Wuling Binguo Ev Datang Lebih Cepat	788	<a href="https://youtu.be/vPfYunGmqfw?si=tLOGHchdcFAirQ2e">https://youtu.be/vPfYunGmqfw?si=tLOGHchdcFAirQ2e</a>
Ridwan Hanif	Alasan Beli Wuling Binguo Ev, Ternyata Begini	571	<a href="https://youtu.be/L1KE49m4KVk?feature=shared">https://youtu.be/L1KE49m4KVk?feature=shared</a>
Pengepul Mobil	Akhirnya Byd M6 Kita Sampai Dirumah !! Apa Aja Kekurangan Dan Kelebihan Nya ??	1.124	<a href="https://youtu.be/xAlOyd1yXZQ?si=z4suhBVxqzddb1os">https://youtu.be/xAlOyd1yXZQ?si=z4suhBVxqzddb1os</a>
Pengepul Mobil	Mobil Listrik Byd M6 Mulai Mengaspal, Yuk Pelajari	511	<a href="https://youtu.be/OI-">https://youtu.be/OI-</a>

	Teknik Marketing "Kuno" Nya !!		<a href="#">fFQjWqTg?feature=shared</a>
Motomobi	BYD M6   MPV Listrik Proper Murah Berfaedah	392	<a href="https://youtu.be/=dvoDG1km9E?s=i=9YsELQN3PJjedQTY">https://youtu.be/=dvoDG1km9E?s=i=9YsELQN3PJjedQTY</a>
<b>Total</b>		<b>4.935</b>	

Data yang akan digunakan dalam format csv yang berisikan teks komentar dari opini masyarakat pada bagian kolom komentar pada video youtube. Kanal youtube yang akan digunakan merupakan kanal yang sudah mendapatkan tanda verifikasi dari youtube yang menandakan bahwa akun tersebut resmi milik seorang kreator atau tokoh publik, sehingga video atau konten-konten yang dihasilkan bersifat orisinal dan dapat lebih dipercaya. Penulis memilih akun-akun tersebut berdasarkan jumlah *viewers* terbanyak dibandingkan konten *creator* lainnya. Berikut merupakan beberapa contoh komentar:

Tabel 3. 2 Contoh komentar pada video ulasan mobil listrik

Kanal video	Akun Pengguna	Waktu	Komentar
Wuling Binguo Ev: Kalau Air Ev Kecelakaan, Beli Ini Aja!	@kirunmv	1 Tahun yang lalu	Gokil 400jtan udah dapet range 400km.... wuling gak main main buat ikut mengubah dunia ke elektrikk car... semoga semakin byk jg splu di indo... Terutama di jalur pantura agr orang mudik ga takut lewat jalur pantura krna kehabisan...
Horee!! Wuling Binguo Ev Datang Lebih Cepat	@ahmadlaibu2739	1 Tahun yang lalu	Belum banyak yg mengevaluasi konsumsi battery ketika menggunakan AC atau ketika AC off, dan

			pengaruhnya terhadap jarak capai maksimum
--	--	--	---

### 3.2 Metode Penelitian

Penelitian ini menerapkan pendekatan kuantitatif dengan tujuan utama untuk mengevaluasi performa algoritma dalam melakukan analisis sentimen terhadap opini publik yang diperoleh dari komentar pada video ulasan mobil di platform YouTube. Pendekatan ini mengacu pada metode sistematis dalam ekstraksi, transformasi, dan analisis data berbasis teks guna memperoleh pemahaman mengenai kecenderungan sentimen masyarakat terhadap isu tertentu. Penelitian ini memanfaatkan beberapa algoritma pembelajaran mesin yang relevan, seperti *Naïve Bayes*, *Support Vector Machine (SVM)*, dan *K-Nearest Neighbor (KNN)*, yang dipilih berdasarkan studi literatur terdahulu mengenai efektivitas model dalam analisis sentimen pada data komentar media sosial, berikut merupakan perbandingan 3 lagoritma yang akan digunakan:

Tabel 3. 3 Perbandingan algoritma yang digunakan

Kategori	Support Vector machine	Naïve Bayes	K-Nearest Neighbor
<b>Akurasi</b>	SVM dikenal menghasilkan akurasi yang sangat tinggi, terutama pada dataset dengan dimensi fitur yang besar dan kompleks. Algoritma ini efektif dalam menemukan hyperplane optimal yang memisahkan kelas secara jelas, sehingga sering mencapai akurasi 95-100% pada berbagai studi, termasuk klasifikasi teks dan data medis.	Naïve Bayes memiliki akurasi yang cukup baik, biasanya berkisar antara 74-87%. Meskipun asumsi independensi fitur jarang terpenuhi secara sempurna, NB tetap memberikan hasil yang stabil dan cukup akurat pada dataset yang tidak terlalu kompleks, seperti data teks sederhana dan klasifikasi spam.	KNN biasanya memberikan akurasi sedang, berkisar antara 75-86%. Algoritma ini bekerja dengan mengklasifikasikan data berdasarkan kedekatan dengan tetangga terdekat, sehingga performanya sangat bergantung pada kualitas data dan pemilihan parameter K. Pada data kecil dan

			bersih, KNN bisa sangat efektif.
<b>Kecepatan Training</b>	Proses training SVM bisa sangat lambat, terutama pada dataset besar dan berdimensi tinggi. Hal ini karena SVM harus mencari hyperplane optimal melalui optimasi yang kompleks dan tuning parameter seperti kernel, C, dan gamma yang memerlukan waktu dan sumber daya komputasi.	Naïve Bayes sangat cepat dalam proses training karena hanya menghitung probabilitas fitur dan kelas tanpa optimasi yang rumit. Kecepatan ini membuat NB sangat cocok untuk aplikasi real-time dan dataset besar yang memerlukan pembelajaran cepat.	KNN tidak memiliki proses training tradisional (lazy learner). Data hanya disimpan, sehingga waktu training hampir nol. Namun, ini berarti beban komputasi dipindahkan ke tahap prediksi.
<b>Kecepatan Prediksi</b>	Setelah training selesai, prediksi dengan SVM relatif cepat karena hanya perlu menghitung posisi data terhadap hyperplane. Namun, pada dataset sangat besar, prediksi bisa menjadi lebih lambat jika model kompleks.	Prediksi Naïve Bayes juga sangat cepat karena hanya melibatkan perhitungan probabilitas sederhana berdasarkan model yang sudah dibangun.	Prediksi KNN cenderung lambat pada dataset besar karena harus menghitung jarak antara data baru dengan seluruh data pelatihan untuk menentukan tetangga terdekat. Ini bisa menjadi bottleneck pada aplikasi dengan data besar.
<b>Kebutuhan Memori</b>	SVM memerlukan memori sedang hingga tinggi, tergantung pada ukuran dataset dan kompleksitas model (jumlah support vectors). Model yang kompleks dengan banyak support vectors akan	Naïve Bayes memiliki kebutuhan memori yang sangat rendah karena hanya menyimpan probabilitas fitur dan kelas, tanpa menyimpan seluruh	KNN membutuhkan memori tinggi karena harus menyimpan seluruh dataset pelatihan agar dapat menghitung jarak saat prediksi. Semakin besar dataset, semakin

	membutuhkan lebih banyak memori.	data pelatihan secara eksplisit.	besar pula kebutuhan memorinya.
<b>Tahan Terhadap Noise</b>	SVM cukup tahan terhadap noise karena berfokus pada support vectors yang paling relevan untuk membentuk hyperplane. Namun, jika data sangat noisy, performa bisa menurun tanpa preprocessing yang baik.	Naïve Bayes kurang tahan terhadap noise karena probabilitas yang dihitung bisa terdistorsi oleh data yang tidak relevan atau outlier, meskipun efeknya tidak terlalu besar karena model sederhana.	KNN sangat sensitif terhadap noise dan outlier karena prediksi bergantung pada tetangga terdekat. Data noisy dapat menyebabkan tetangga yang salah dipilih dan menurunkan akurasi.
<b>Kebijakan</b>	SVM unggul dalam menghasilkan akurasi tinggi pada data berdimensi tinggi dan kompleks, serta memiliki kemampuan generalisasi yang baik jika parameter di-tuning dengan benar.	Naïve Bayes sangat cepat, mudah diimplementasikan, dan efisien untuk data teks dan aplikasi real-time. Cocok untuk dataset besar dengan fitur independen.	KNN mudah dipahami dan diimplementasikan, tidak memerlukan proses training, dan efektif untuk dataset kecil dengan distribusi data yang jelas. Baik untuk kasus di mana recall sangat penting.
<b>Kekurangan</b>	Training SVM bisa lambat dan membutuhkan tuning parameter yang tepat. Tidak cocok untuk dataset yang sangat besar tanpa optimasi khusus.	Asumsi independensi fitur jarang terpenuhi di dunia nyata, sehingga akurasi NB bisa terbatas pada dataset kompleks.	Prediksi lambat pada dataset besar, sangat sensitif terhadap noise, dan performa menurun jika data tidak dinormalisasi atau fitur tidak relevan.
<b>Bagus digunakan untuk</b>	Data teks, data medis, dan dataset berdimensi tinggi yang memerlukan akurasi tinggi dan generalisasi baik.	Aplikasi <i>real-time</i> , <i>spam filtering</i> , <i>sentiment analysis</i> , dan dataset besar dengan fitur independen.	Dataset kecil, aplikasi di mana recall sangat penting, dan data yang relatif bersih dan

			terstruktur dengan baik.
--	--	--	--------------------------

Pemilihan algoritma *Naïve Bayes* (NB), *Support Vector Machine* (SVM), dan *K-Nearest Neighbor* (KNN) untuk analisis sentimen pada komentar video review mobil listrik di YouTube didasarkan pada karakteristik dataset yang digunakan serta kemampuan masing-masing algoritma dalam memberikan hasil yang optimal. *Naïve Bayes* dipilih karena kecepatan dan efisiensinya dalam mengolah data teks dalam jumlah besar, sehingga sangat cocok untuk pemrosesan cepat komentar yang banyak. SVM digunakan karena kemampuannya dalam menangani data berdimensi tinggi dan kompleks, memberikan akurasi yang tinggi dalam klasifikasi sentimen yang bervariasi. Sedangkan KNN dipilih karena kesederhanaannya dan kemampuannya dalam mengenali pola berdasarkan kemiripan data, yang membantu meningkatkan recall pada beberapa kasus. Kombinasi ketiga algoritma ini memungkinkan analisis sentimen yang lebih komprehensif dan optimal sesuai dengan karakteristik data komentar video YouTube.

### 3.2.1 Alur Penelitian

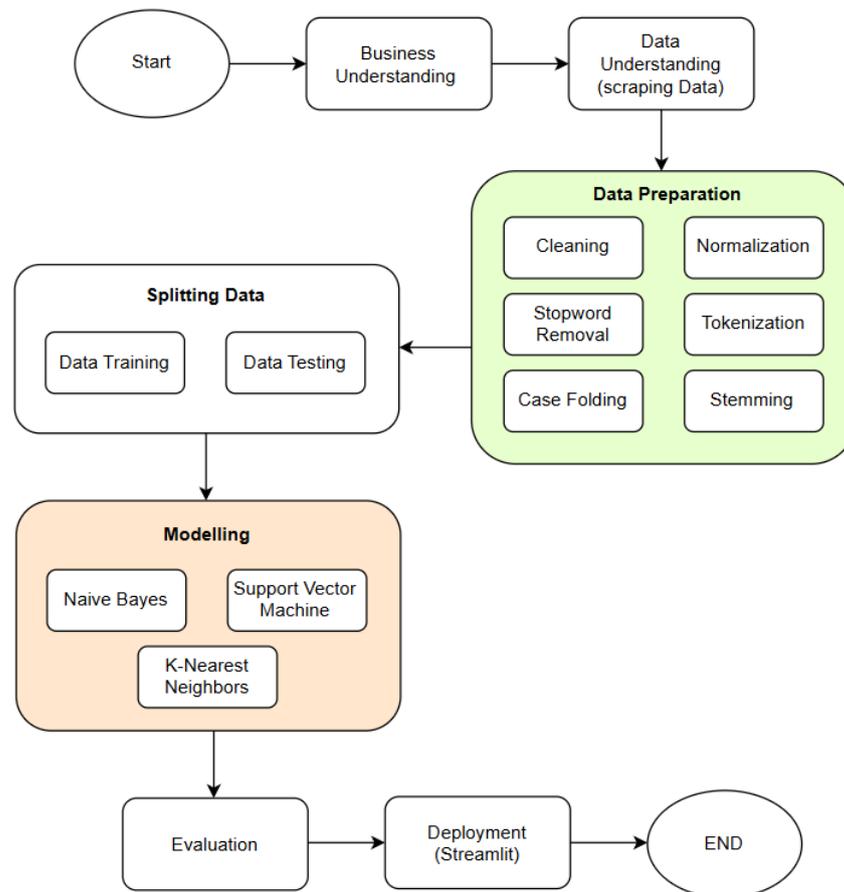
Teknik analisis data dalam penelitian ini menggunakan metodologi *Cross Industry Standard Process for Data Mining* (CRISP-DM), yang merupakan kerangka kerja standar dalam bidang penambangan data. Metodologi ini terdiri dari enam tahapan utama, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Setiap tahapan tersebut dirancang secara sistematis untuk memastikan proses analisis data berjalan terstruktur dan menghasilkan model yang sesuai dengan tujuan penelitian.

Tabel 3. 4 Perbandingan crisp-dm dan kdd

CRISP-DM	KDD
Memiliki enam tahapan: <i>business understanding</i> , <i>data understanding</i> , <i>data preparation</i> , <i>modeling</i> , <i>evaluation</i> , dan <i>deployment</i> .	Memiliki 5 tahapan pengembangan data mining yaitu <i>selection</i> , <i>pre processing</i> , <i>transformation</i> , <i>data mining</i> , <i>interpretation/evaluation</i> .
Mencakup tahap akhir berupa <i>deployment</i> , yaitu implementasi model ke lingkungan nyata.	Tidak memiliki tahapan <i>deployment</i> sebagai bagian dari proses.

Dimulai dengan pemahaman tujuan bisnis dan analisis kebutuhan data ( <i>business &amp; data understanding</i> ).	Dimulai dengan proses seleksi data sebagai dasar pembuatan dataset.
--	---

Pada penelitian ini, alur penelitian yang dilakukan akan berjalan dengan alur seperti berikut:



Gambar 3.1 Alur penelitian

Gambar 3.1 menyajikan diagram alur yang menggambarkan secara sistematis rangkaian proses yang diterapkan dalam penelitian ini. Diagram tersebut dirancang dengan mengadopsi metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai kerangka kerja yang menjadi panduan dalam pelaksanaan setiap tahapan. Berdasarkan pendekatan CRISP-DM, terdapat enam fase utama yang diimplementasikan secara berurutan dan iteratif dalam penelitian ini, yaitu:

## **1. Business Understanding**

Pada tahap ini, fokus utama adalah memahami permasalahan yang menjadi dasar dilakukannya penelitian serta menetapkan tujuan yang ingin dicapai. Penelitian ini bertujuan untuk menganalisis dampak sentimen publik terhadap komentar-komentar yang muncul pada video review mobil listrik di YouTube. Untuk mendapatkan gambaran yang komprehensif mengenai persepsi masyarakat, analisis sentimen dilakukan dengan menggunakan data komentar publik yang dikumpulkan sejak mobil listrik tersebut diluncurkan atau sejak konten review mulai dibuat.

Pendekatan ini memungkinkan peneliti untuk melihat bagaimana sentimen publik berkembang dari awal peluncuran hingga periode berikutnya, sehingga dapat mengidentifikasi perubahan sikap dan opini masyarakat terhadap mobil listrik. Hasil dari analisis ini diharapkan dapat memberikan manfaat bagi berbagai pihak terkait, seperti perusahaan otomotif Wuling dan BYD, *dealer* mobil listrik, serta para peneliti, dalam mengambil keputusan yang lebih tepat berdasarkan persepsi pengguna. Selain itu, dengan melakukan perbandingan kinerja antara algoritma NB, SVM, dan KNN, penelitian ini bertujuan untuk menentukan metode analisis sentimen yang paling efektif dalam mengklasifikasikan opini masyarakat mengenai mobil listrik.

## **2. Data Understanding**

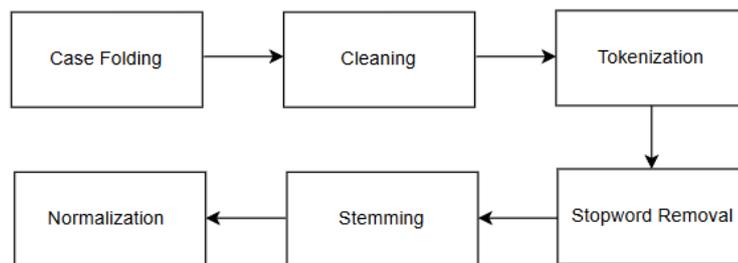
Pada tahap ini, data dikumpulkan dengan menggunakan kata kunci yang relevan, yaitu "BYD M6" dan "Wuling Cloud EV". Pemilihan *video review* dilakukan dengan menyaring berdasarkan jumlah penonton (*view count*) agar memperoleh video dengan jangkauan penonton terbanyak. Video yang diambil berasal dari *channel* yang sudah terverifikasi oleh YouTube. Tidak berasal dari perusahaan resmi atau video yang bersifat iklan, sehingga data yang diperoleh lebih objektif dan akurat. Dataset yang terkumpul terdiri dari sekitar 4.935 komentar yang menjadi sumber utama

dalam analisis sentimen terhadap opini publik mengenai mobil listrik BYD M6 dan Wuling Cloud EV. Pengambilan data dilakukan melalui proses *scraping* menggunakan YouTube API dengan bahasa pemrograman Python, dan seluruh proses ini dijalankan di lingkungan Google Colab untuk memudahkan pengelolaan dan pemrosesan data.

### 3. *Data Preparation*

Tahapan berikutnya adalah mempersiapkan data yang akan digunakan sesuai dengan kebutuhan dalam pembuatan model.

Dengan alur sebagai berikut:



Gambar 3.2 Alur proses *data preparation*

1. *Case Folding*, merupakan proses konversi seluruh huruf dalam teks menjadi huruf kecil. Tahapan ini bertujuan untuk mengurangi redundansi kata akibat perbedaan kapitalisasi, seperti kata "Data" dan "data" yang seharusnya dianggap sama. Dengan menyamakan bentuk huruf, proses analisis teks selanjutnya menjadi lebih konsisten dan akurat.
2. *Cleaning*, tahap pembersihan data teks dari elemen-elemen yang tidak relevan atau mengganggu proses analisis. Proses ini mencakup penghapusan tautan (URL), mention (@username), hashtag (#tag), karakter non-alfabet seperti simbol atau angka, tanda baca, serta spasi yang berlebihan. Tujuan utama dari pembersihan ini adalah untuk menghilangkan noise atau gangguan dalam data teks agar proses selanjutnya tidak

terpengaruh oleh informasi yang tidak relevan atau mengganggu. Hal ini penting karena data teks yang diambil dari internet, media sosial, atau komentar pengguna biasanya memiliki banyak elemen yang tidak diperlukan dalam konteks analisis.

3. *Tokenizing*, proses memecah teks atau kalimat menjadi unit-unit yang lebih kecil yang disebut token, yang umumnya berupa kata. Proses ini penting karena memungkinkan analisis teks dilakukan pada level kata, bukan pada level kalimat atau paragraf. Tokenisasi membantu dalam mengenali struktur dan makna kata dalam kalimat, serta memudahkan dalam penerapan stemming, stopword removal, dan penghitungan frekuensi kata. Tujuan utamanya adalah untuk mengorganisasi teks mentah menjadi bentuk yang lebih terstruktur sehingga dapat dianalisis secara matematis atau statistik oleh model pembelajaran mesin.
4. *Stopword Removal*, proses menghapus kata-kata umum atau kata bantu dalam bahasa tertentu yang sering muncul namun tidak memiliki kontribusi makna signifikan terhadap analisis. Contoh kata-kata ini dalam bahasa Indonesia adalah “dan”, “di”, “ke”, “adalah”, dan sebagainya. Meskipun kata-kata ini penting dalam pembentukan kalimat secara tata bahasa, dalam konteks analisis teks, kehadirannya justru bisa mengganggu dan membuat model menjadi bias. Dengan menghapus stopword, kita dapat meningkatkan kualitas fitur yang digunakan, serta mengurangi kompleksitas data dan mempercepat proses analisis.
5. *Stemming*, proses mengubah kata ke bentuk dasarnya atau akarnya dengan cara menghapus imbuhan, seperti awalan, akhiran, atau sisipan. Sebagai contoh, kata "bermain", "permainan", dan "bermain-main" akan distem menjadi kata dasar "main". Proses ini sangat penting untuk mengurangi

variasi bentuk kata yang berbeda tetapi memiliki makna yang sama, sehingga membantu menyederhanakan representasi kata. Tujuan stemming adalah untuk meningkatkan efisiensi dalam pencocokan kata dan analisis frekuensi kata, serta memperkecil dimensi fitur dalam model machine learning.

6. *Normalization*, tahapan untuk menyamakan bentuk penulisan kata agar menjadi lebih standar dan konsisten. Hal ini mencakup penghilangan karakter berulang seperti pada kata “hebaaattt” menjadi “hebat”, serta mengoreksi bentuk tidak baku atau informal menjadi bentuk baku. Dalam konteks bahasa media sosial atau komentar online, kata-kata sering kali ditulis dengan cara tidak standar. Oleh karena itu, normalisasi menjadi sangat penting untuk memastikan bahwa kata-kata tersebut bisa dikenali dan diproses dengan benar oleh model analisis. Tujuan normalisasi adalah untuk meningkatkan keakuratan analisis dengan menyederhanakan variasi penulisan kata menjadi bentuk yang konsisten.

#### 4. *Modelling*

Pada tahap ini, dilakukan beberapa proses terlebih dahulu seperti pelabelan data menggunakan TextBlob, transformasi teks dengan metode TF-IDF, dan pembagian data (data splitting). TextBlob digunakan karena telah memiliki kamus sentimen yang memudahkan klasifikasi opini menjadi positif, negatif, atau netral secara otomatis. Selanjutnya, TF-IDF digunakan untuk mengubah teks menjadi representasi numerik berdasarkan frekuensi kemunculan kata, sehingga teks dapat diolah oleh algoritma pembelajaran mesin.

Dalam menangani ketidakseimbangan jumlah data antar kelas, diterapkan metode SMOTE dalam penelitian yang menghasilkan data sintetis dari kelas minoritas agar model dapat belajar secara seimbang dan menghasilkan klasifikasi yang lebih

akurat. Sebelum proses pemodelan dimulai, data akan dibagi terlebih dahulu melalui teknik *splitting*. Data ulasan akan dipisahkan menjadi dua bagian, yaitu 80% data digunakan untuk proses pelatihan model (*training*), dan 20% sisanya digunakan untuk menguji kinerja model (*testing*).

Setelah data dipersiapkan melalui tahapan tersebut, proses pemodelan dilakukan menggunakan tiga algoritma klasifikasi, yaitu *Naive Bayes* (NB), *Support Vector Machine* (SVM), dan *K-Nearest Neighbor* (KNN) untuk mengevaluasi performa prediksi sentimen secara komparatif, kemudian hasilnya akan dibandingkan untuk mengetahui model mana yang terbaik dengan melakukan evaluasi.

#### **5. Evaluation**

Tahap berikutnya adalah melakukan evaluasi terhadap model yang sudah dibuat dengan tiga algoritma yang berbeda. Model yang telah diimplementasikan akan dievaluasi dengan *matrix confusion* untuk mendapatkan hasil akurasi, *precision*, *recall*, dan *F1-score*. Evaluasi ini dilakukan untuk membandingkan performa ketiga algoritma dalam melakukan analisis sentimen review. Hasil evaluasi diperlihatkan dalam bentuk tabel dan grafik guna memudahkan interpretasi dan analisis. Evaluasi juga akan mencakup topik-topik utama yang dibahas dalam tiga kategori sentimen, yaitu positif, negatif, dan netral.

#### **6. Deployment**

Setelah proses evaluasi model selesai dan diperoleh algoritma terbaik berdasarkan performanya, tahap selanjutnya adalah penerapan hasil analisis ke dalam bentuk aplikasi berbasis web. Dalam studi ini, pengembangan aplikasi dilakukan menggunakan Streamlit, yaitu sebuah framework Python yang

sederhana dan efisien untuk membangun antarmuka pengguna secara interaktif. Streamlit berfungsi sebagai dashboard analisis sentimen, di mana pengguna dapat dengan mudah melihat hasil klasifikasi sentimen dari data yang telah dilabeli, baik dalam bentuk visualisasi maupun teks. Dengan adanya Streamlit, implementasi model analisis sentimen menjadi lebih praktis dan dapat diakses oleh pengguna non-teknis. Hal ini memungkinkan pengguna untuk memanfaatkan hasil analisis secara langsung dalam pengambilan keputusan, tanpa harus memahami proses teknis di balik pembangunan model.

### 3.3 Teknik Pengumpulan Data

Dalam penelitian ini, data yang dianalisis diperoleh melalui proses *scraping* dari platform media sosial YouTube. Scraping sendiri merupakan metode pengambilan data secara otomatis dari halaman web, dalam hal ini digunakan untuk mengumpulkan komentar dari enam video yang berbeda. Proses ini dilakukan menggunakan bantuan perangkat lunak (*software*) yang memungkinkan komputer untuk mengakses dan mengekstrak data secara efisien tanpa harus dilakukan secara manual.

YouTube, melalui layanan *Application Programming Interface* (API)-nya, menyediakan akses terbuka yang dapat dimanfaatkan untuk kebutuhan penelitian, termasuk *scraping* data. YouTube Data API v3 dari Google Cloud memungkinkan pengguna untuk memperoleh berbagai informasi dari konten video secara sistematis dan legal. Dalam penelitian ini, proses *scraping* diimplementasikan menggunakan bahasa pemrograman Python yang dikenal fleksibel dan didukung berbagai pustaka (*library*) yang sangat membantu dalam melakukan pengambilan data, seperti *google-api-python-client*, *pandas*, dan lainnya.

Pengambilan data komentar dilakukan dengan menggunakan dua kata kunci, yaitu 'Wuling Binguo EV' dan 'BYD M6'. Kemudian, data yang telah didapatkan akan difilter dan diurutkan berdasarkan penonton terbanyak dengan

memilih opsi ‘view count’ pada menu filter. dilakukan pada bulan Februari 2025 dan menghasilkan total 4.935 entri komentar. Semua data yang berhasil dikumpulkan kemudian disimpan dalam format .csv (*Comma Separated Values*), yang memudahkan proses pemrosesan data lebih lanjut dalam tahap analisis dan modeling. Penyimpanan dalam format ini juga mempermudah integrasi dengan berbagai alat analisis data lainnya. Dengan demikian, proses *scraping* menjadi langkah awal yang krusial dalam membentuk fondasi dataset untuk analisis sentimen yang dilakukan dalam studi ini.

### **3.4 Variabel Penelitian**

#### **3.4.1 Variabel Independen**

Variabel Independen merupakan variabel yang menjadi penyebab atau terpengaruhnya variabel terikat. Berikut merupakan variabel independen pada penelitian ini :

1. Komentar dari video ‘Wuling Binguo Ev: Kalau Air Ev Kekecilan, Beli Ini Aja!’ pada kanal Youtube Ridwan Hanif.
2. Komentar dari video ‘Horee!! Wuling Binguo Ev Datang Lebih Cepat’ pada kanal Youtube Fitra Eri.
3. Komentar dari video ‘Alasan Beli Wuling Binguo Ev, Ternyata Begini’ pada kanal Youtube Ridwan Hanif.
4. Komentar dari video ‘Akhirnya Byd M6 Kita Sampai Dirumah !! Apa Aja Kekurangan Dan Kelebihan Nya ??’ pada kanal Youtube Pengepul Mobil.
5. Komentar dari video ‘Mobil Listrik Byd M6 Mulai Mengaspal, Yuk Pelajari Teknik Marketing "Kuno" Nya !!’ pada kanal Youtube Pengepul Mobil.
6. Komentar dari video ‘BYD M6 | MPV Listrik Proper Murah Berfaedah’ pada kanal Youtube Motomobi.

#### **3.4.2 Variabel Dependen**

Variabel Dependen merupakan variabel terikat yang dipengaruhi karena adanya variabel independen [10]. Variabel Dependen pada

penelitian ini merupakan tanggapan yang bersifat positif, netral dan negatif sehingga terdapat 3 kelas klasifikasi yaitu positif, netral dan negatif. Variabel tersebut bergantung kepada setiap variabel independen yaitu masing – masing tanggapan dari komentar Youtube pada konten yang dibuat oleh akun Ridwan Hanif, Pengepul Mobil, Motomobi, dan Fitra Eri.

### 3.5 Teknik Analisis Data

*Tools* yang digunakan dalam penelitian ini adalah bahasa pemrograman Python, yang memiliki sejumlah keunggulan seperti kecepatan dalam komputasi, dukungan teknologi canggih yang memungkinkan pengolahan dan analisis data dalam skala besar, serta kemudahan penggunaan yang membuatnya sangat ramah bagi peneliti. Python dipilih karena kemampuannya dalam mendukung proses *text mining* dan analisis data yang kompleks, sekaligus mudah dipelajari dan memiliki komunitas pengguna yang luas.

Untuk mendukung proses analisis sentimen dalam penelitian ini, digunakan *Integrated Development Environment* (IDE) yang populer dan berbasis *cloud*, yaitu Google Colab. Google Colab menjadi pilihan karena menawarkan berbagai keunggulan, seperti akses gratis ke komputasi berbasis GPU, integrasi langsung dengan Google Drive untuk penyimpanan data, serta kemudahan berbagi dan kolaborasi secara daring. Selain itu, antarmuka Google Colab menyerupai Jupyter Notebook sehingga tetap mendukung visualisasi output secara interaktif, tetapi dengan tambahan fleksibilitas dalam akses dari berbagai perangkat tanpa perlu instalasi lokal. Dengan demikian, Google Colab dipilih sebagai platform pengembangan Python yang efisien, fleksibel, dan praktis untuk mendukung keseluruhan tahapan analisis dalam penelitian ini.