

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Objek penelitian ini adalah untuk menganalisis sentiment masyarakat Indonesia terhadap mobil listrik. Mobil listrik dipercaya dapat berkontribusi untuk mengurangi emisi CO₂ dari perjalanan, terutama ketika produksi dan penggunaannya ditenagai dengan energi terbarukan. Mobil ramah lingkungan sedang berkembang saat ini, dengan mobil listrik sebagai salah satu solusi yang dipilih, terutama di Indonesia, yang merupakan negara dengan jumlah penduduk terbesar keempat di dunia. Dengan berbagai isu tentang mobil listrik mencakup masalah ketersediaan infrastruktur pengisian daya, keterbatasan jarak tempuh, dan biaya awal yang tinggi. Selain itu, keberlanjutan energi dan dampak lingkungan dari pembuatan baterai juga menjadi perhatian. Indonesia sedang berusaha untuk mengadopsi teknologi mobil listrik sebagai bagian dari transportasi umum. Analisis sentimen ini dilakukan untuk memberikan pemahaman menyeluruh mengenai informasi tingkat akhir yang ditinggalkan pengguna pada halaman aplikasi twitter. Data dari Twitter digunakan peneliti untuk menentukan metode klasifikasi yang dapat memberikan tingkat akurasi dan berfokus pada sentiment positif dan negatif mobil listrik.

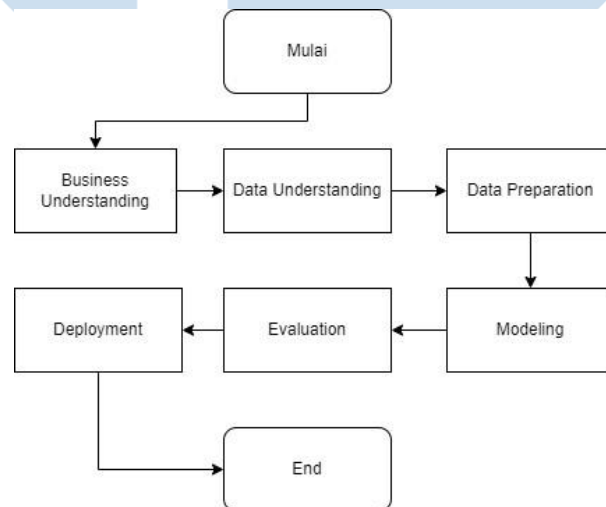
3.2 Metode Penelitian

Penelitian ini menggunakan metode penelitian kualitatif, dimana akan dilakukan pengumpulan data dari aplikasi twitter atau X. Metode Kualitatif diterapkan dengan mengimplementasikan scraping data yang memperoleh data tweet mobil listrik pada masyarakat diindonesia. Alasan di balik pemilihan Naive Bayes karena terkenal karena kesederhanaan dan efisiensinya, serta kemampuannya dapat memberikan hasil yang baik bahkan dengan dataset yang relatif kecil, selain itu algoritma ini bekerja efektif untuk teks dengan ribuan kata unik [44] [45], Pada penelitian terdahulu pembagian data 80:20, mencapai akurasi sebesar 81%, namun secara keseluruhan, model menunjukkan kinerja

yang relatif stabil dengan nilai akurasi yang cukup tinggi, setelah itu ada pelabelan menggunakan vader lexicon pemilihan vader lexicon karena lebih populer khususnya untuk sentiment analisis yang dikenal karena kehandalannya dalam pelabelan, keuntungan menggunakan vader lexicon karena tersedianya kamus dengan memuat berbagai nilai pada masing-masing kata, vader lexicon juga memiliki keunggulan berupa kata dalam sebuah kalimat yang akan dibandingkan secara langsung dengan kamus opini, jika di dalam kalimat tersebut tersedia kata yang bersifat opini, maka kalimat tersebut akan disebut kalimat opini. Dan jika kata yang tersedia dalam kalimat tersebut tidak tersedia dalam kamus Lexicon, maka dianggap bukan kalimat opini [55].

3.2.1 Alur Penelitian

Dalam alur penelitian ini diterapkan beberapa langkah-langkah yang digambarkan pada Gambar 3.1.



Gambar 3.1 Alur penelitian

Berikut ini penjelasan dari tahapan yang akan digunakan dalam penelitian ini, antara lain:

3.2.1.1 Business Understanding

Pada tahap pertama ini dilakukan teknik analisis menggunakan metode CRISP-DM tahap ini bertujuan untuk memahami penelitian yang mendalam mengenai bidang bisnis, termasuk metode pengumpulan data

yang digunakan untuk mengembangkan strategi bisnis dari informasi yang diperoleh. Penelitian ini difokuskan untuk memberikan wawasan yang berfokus pada pembahasan isu masalah mobil listrik yang terjadi pada masyarakat Indonesia yang dapat ditinjau berdasarkan komentar masyarakat Indonesia melalui data aplikasi twitter.

3.2.1.2 Data Understanding

Penelitian ini menggunakan data ulasan 3053 dari Twitter yang dikumpulkan berdasarkan pandangan masyarakat Indonesia terhadap mobil listrik melalui alat scrapping Google Colab menunjukkan upaya yang dalam pengumpulan data. Melalui penggunaan data scrapping, dengan begitu dapat mengakses informasi dari berbagai sumber dan mengumpulkan ulasan yang mencerminkan pandangan yang beragam dari masyarakat. Atribut yang dikumpulkan, seperti tanggal pembuatan, jumlah komentar, jumlah balasan, jumlah retweet, jumlah favorit, dan nama pengguna, memberikan informasi terkait dengan interaksi pengguna dengan topik yang diteliti. Dengan jumlah data yang signifikan ini, analisis yang mendalam tentang sentimen dan opini masyarakat terhadap mobil listrik dapat dilakukan dengan tingkat harapan statistik yang lebih tinggi.

3.2.1.3 Data Preparation

Pada tahap ini, tahapan penting dalam analisis data yang bertujuan untuk membersihkan, mempersiapkan, dan mengorganisir dengan proses seleksi data berdasarkan komentar, dan tanggal komentar. Setelah itu data melewati tahapan-tahapan dalam data *preprocessing* meliputi *cleaning*, *case folding*, *tokenization*, dan *stemming data* yang akan dilakukan. Berikut ini merupakan penjelasan dari Cara data preparation yang berurutan mulai dari tahap awal hingga akhir:

1. *Cleaning*

```

import re

def clean_twitter_text(text):
    # Menghapus nama pengguna Twitter (@username)
    text = re.sub(r'@[A-Za-z0-9_]+', '', text)

    # Menghapus hashtag (#kata)
    text = re.sub(r'#\w+', '', text)

    # Menghapus 'RT' (retweet) dan spasi yang mengikutinya
    text = re.sub(r'RT[\s]+', '', text)

    # Menghapus URL (http:// atau https:// diikuti oleh karakter non-whitespace)
    text = re.sub(r'https?://\S+', '', text)

    # Menghapus karakter selain huruf, angka, atau spasi
    text = re.sub(r'^A-Za-z0-9\s$', '', text)

    # Menghapus spasi ekstra dan membersihkan teks
    text = re.sub(r'\s+', ' ', text).strip()

    return text

# Menggunakan fungsi clean_twitter_text untuk membersihkan kolom 'full_text' di DataFrame
data['full_text'] = data['full_text'].apply(clean_twitter_text)

data.to_csv('/content/data_mobil_listrik.csv', sep=";", index=False)

```

Gambar 3.2 Cara *Cleaning*

Cara ini dilakukan untuk menghapus satu huruf, URL, dan karakter non-huruf (seperti tanda baca dan angka) dari data teks. Hal ini dilakukan karena karakter tunggal, tanda baca, URL, dan karakter non-huruf dianggap sebagai elemen yang tidak berguna dan tidak penting. Item-item ini, jika tidak dicentang, akan diubah menjadi karakteristik pemodelan yang mempengaruhi hasil evaluasi kinerja model.

2. *Case Folding*

```

def case_folding(text):
    if isinstance(text, str):
        lowercase_text = text.lower()
        return lowercase_text
    else:
        return text

df['case_folding'] = df['cleaning'].apply(case_folding)
df.head(5)

```

Gambar 3.3 Cara *Case Folding*

Proses ini bertujuan adalah untuk mengonversi semua karakter menjadi huruf kecil. Sebagai contoh mengubah "Mobil" menjadi

"mobil". Tidak hanya itu, karakter non-huruf akan dihapus dan seringkali digunakan sebagai pemisah antara kata-kata.

3. Tokenizing

```
def tokenize(text):  
    token = text.split()  
    return tokens  
  
df['tokenize'] = df['case_folding'].apply(tokenize)  
df.head(5)
```

Gambar 3.4 Cara Tokenization

Pada bagian ini, teks diberi token yaitu, dibagi menjadi segmen atau atribut token. Teks dibagi menjadi beberapa bagian sesuai dengan ruang yang digunakan untuk tokenisasi dalam penelitian ini. Untuk menyederhanakan.

4. Stemming Data

```
factory = StemmerFactory()  
stemmer = factory.create_stemmer()  
def stem_text(text):  
    return [stemmer.stem(word) for word in text]  
  
df['stemming_data'] = df['filtering/stopword removal'].apply(lambda x: ' '.join(stem_text(x)))  
df.head(5)
```

Gambar 3.5 Cara stemming data

Kata-kata diubah menjadi bentuk paling dasar pada saat ini. Hal ini perlu dilakukan karena kata-kata yang mempunyai imbuhan yang berasal dari kata dasar yang sama diperlakukan sebagai kata-kata tersendiri. Di sinilah dilakukan proses penggabungan berbagai bentuk kata menjadi satu representasi.

3.2.1.4 Modeling

Pada tahap pemodelan, penelitian ini mengimplementasikan algoritma analisis sentimen, yaitu Naïve Bayes pada dataset ulasan aplikasi twitter yang telah melalui tahap pra-pemrosesan. Pertama, dilakukan pembagian dataset menjadi dua bagian, yaitu data latih dan data uji. Data latih digunakan untuk melatih model, sementara data uji

digunakan untuk menguji kinerja model. Kemudian, algoritma Naïve Bayes diterapkan untuk pembuatan model, dengan menggunakan data latih sebagai bahan pelatihan. Evaluasi dilakukan dengan menggunakan data uji, dan metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score* dihitung untuk mengevaluasi kinerja model Naïve Bayes.

3.2.1.5 Evaluation

Evaluasi merupakan langkah dari proses yang telah dilakukan pada tahap sebelumnya, yaitu pemodelan. Evaluasi lebih lanjut dilakukan untuk menentukan apakah hasil dari pemodelan tersebut memenuhi tujuan penelitian atau tidak. Pada tahap evaluasi ini, akan dihasilkan evaluasi dari model naïve bayes, selain itu ada analisis sentimen berupa hasil positif, dan negatif. dan yang terakhir juga akan dibuat *wordcloud* yang menampilkan kata-kata yang sering muncul dalam ulasan.

3.2.1.6 Deployment

Pada tahap deployment dilakukan implementasi model naïve bayes yang telah dilakukan sebelumnya ditahapan *evaluation*, dengan digunakan implementasi tersebut bisa digunakan untuk menghasilkan skor keputusan akhir yang tepat dan memenuhi tujuan penelitian, serta menghasilkan visualisasi data.

3.3 Teknik Pengumpulan Data

Tahap pengumpulan data dalam penelitian ini dilakukan melalui dua proses utama yaitu pengumpulan data dari tweet pengguna aplikasi Twitter dan menerapkan teknik perayapan data menggunakan Google Colab dengan perpustakaan google-play-scraper. Berikut penjelasan pengumpulan datanya tahapan:

1. Identifikasi Sumber Data:

Sumber data utama yang diidentifikasi adalah Respon Masyarakat Indonesia terhadap mobil Listrik dihalaman twitter. Pengumpulan data dilakukan fokus pada respon pengguna, termasuk informasi seperti nama pengguna, tanggal, dan tweet.

2. Jumlah Data yang Diambil:

Data dikumpulkan dari sepuluh ribu terbaru tweet pengguna pada Juni – Januari 2024. Jumlah ini dipilih agar mendapatkan jumlah yang cukup besar dan representative.

3. Teknik Pengumpulan Data:

Pengumpulan data menggunakan teknik data *crawling* yaitu pengambilan data secara otomatis dari halaman aplikasi twitter. Alat yang digunakan untuk perayapan data adalah google colab, platform berbasis *cloud* untuk pemrograman python, memanfaatkan perpustakaan google-play-scraper. Pustaka ini digunakan di lingkungan Google Colab untuk memfasilitasi proses mengambil data dari Twitter. Dengan perpustakaan ini, proses *scraping* menjadi lebih efisien dan tersusun.

4. Variabel Data yang Diambil:

Data yang diambil mencakup empat variabel utama: nama pengguna Tanggal, dan Tweet. Variabel-variabel ini dipilih karena memberikan arti penting informasi terkait sentimen pengguna terhadap Mobil listrik.

5. Proses Pengumpulan Data:

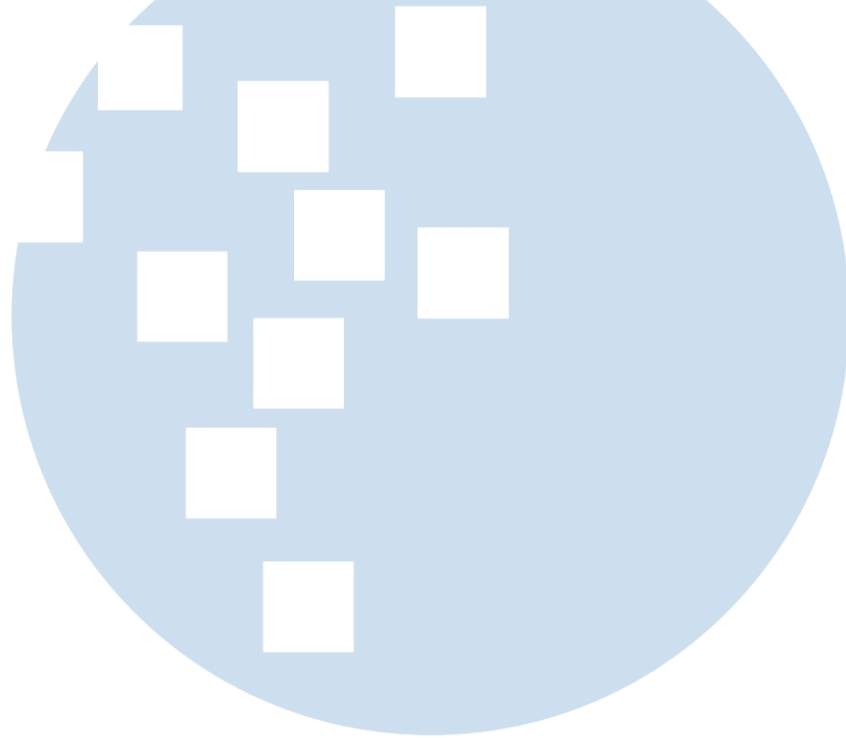
Proses pengumpulan data dilakukan dengan mengakses Halaman Twitter, lalu ekstrak yang diperlukan informasi dari setiap ulasan pengguna. Pustaka google-play-scraper membantu mengelola pengikisan proses dengan menyediakan akses yang terstruktur dan efisien untuk meninjau data.

6. Penggunaan Teknik Crawling:

Teknik *crawling* digunakan untuk menjelajahi Twitter halaman dan secara sistematis mengumpulkan ulasan pengguna informasi. Proses ini dilakukan dengan memanfaatkan kerohan kemampuan Google Colab dan perpustakaan google-play-scraper.

Melalui pendekatan ini, peneliti dapat memastikan bahwa informasi yang dikumpulkan adalah informasi terkini, relevan, dan mencakup rentang

data tingkat respon Masyarakat terhadap mobil listrik pada waktu tertentu. Hal ini berfungsi sebagai landasan penting untuk menyiapkan data untuk tahap analisis signifikansi berikutnya.



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA