

BAB 3

METODE PENELITIAN

Penelitian ini merupakan analisis persepsi masyarakat untuk mengidentifikasi sentimen dan topik terhadap penggunaan motor listrik pada ojek online menggunakan metode SVM dan LDA. Berikut ini merupakan bahan, alat, dan jalan penelitian dalam melakukan penelitian ini.

3.1 BAHAN DAN ALAT PENELITIAN

Bahan penelitian dalam penelitian merupakan data komentar TikTok dari beberapa video yang berkaitan dengan ojek online dengan motor listrik. Data tersebut diambil menggunakan platform Apify dan *scraping* manual menggunakan API TikTok.

Alat yang digunakan dalam penelitian ini adalah komputer dengan spesifikasi cukup untuk menjalankan sistem operasi dan perangkat lunak pengembangan serta koneksi Internet.

Sistem Operasi dan program-program aplikasi yang dipergunakan dalam penelitian ini adalah:

1. Sistem Operasi Windows 11.
2. Bahasa pemrograman Python.
3. Google Colab.

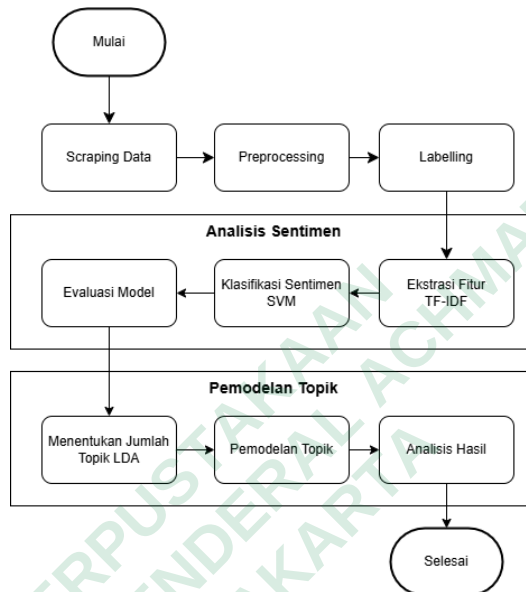
Library yang digunakan dalam pengolahan data pada penelitian ini adalah sebagai berikut:

1. Pandas versi 2.2.2
2. NumPy versi 2.0.2
3. NLTK versi 3.9.1
4. Scikit-learn versi 1.6.1
5. Sastrawi versi 1.0.1
6. TextBlob versi 0.19.0
7. Joblib versi 1.5.1
8. Matplotlib versi 3.10.0

9. Seaborn versi 0.13.2
10. Gensim versi 4.3.3
11. pyLDAvis versi 3.4.1

3.2 JALAN PENELITIAN

Jalan penelitian dalam ini penelitian dapat dilihat pada Gambar 3.1.



Gambar 3.1 Jalan Penelitian

Berikut merupakan langkah-langkah jalannya penelitian yang dilakukan, yaitu:

1. *Scraping Data*

Pengumpulan data dilakukan dengan menggunakan platform Apify dan *scraping* manual menggunakan API TikTok. Data yang digunakan merupakan data komentar TikTok dari video yang membahas tentang motor listrik pada ojek online. Proses pengambilan data dilakukan dengan memasukkan kata kunci seperti "ojol motor listrik", "gojek electric", dan "pov naik ojol listrik", kemudian memilih video yang relevan dan beragam komentarnya.

2. *Preprocessing*

Pada tahap ini data yang sudah diperoleh akan dibersihkan agar data siap untuk dianalisis. Preprocessing data terdiri dari beberapa tahap, yakni (Riyandona et al., 2025):

a. *Cleaning*

membersihkan simbol, link, hashtag dan lainnya yang tidak memiliki arti.

b. *Case Folding*

Merubah huruf kapital menjadi huruf kecil semua.

c. *Normalisasi*

Mengubah kata tidak baku atau singkatan menjadi baku.

d. *Tokenization*

Memecah kalimat menjadi sebuah pecahan-pecahan kata.

e. *Stopword Removal*

Menghapus kata-kata yang tidak berguna dan tidak memiliki arti, seperti "di", "ke", "yang".

f. *Stemming*

Merubah sebuah kata menjadi kata dasar.

3. *Labelling*

Memberikan label pada data yang sudah dibersihkan ke dalam beberapa kategori sentimen, yaitu positif, negatif, dan netral. Sentimen positif untuk komentar yang mendukung terhadap penggunaan motor listrik pada ojek online. Sebaliknya, sentimen negatif untuk komentar yang berisi kritikan atau pernyataan yang kurang mendukung. Sementara label netral digunakan untuk komentar yang tidak dapat diidentifikasi menjadi komentar positif maupun negatif.

4. **Ekstrasi Fitur TF-IDF**

Memberi bobot pada sebuah kata untuk melihat seberapa penting suatu kata dalam komentar. Kata yang sering muncul dalam satu komentar tapi jarang muncul di komentar lain akan memiliki bobot yang tinggi. Kata umum yang sering muncul di semua komentar akan memiliki bobot lebih rendah.

5. **Klasifikasi Sentimen SVM**

Menentukan apakah komentar bersifat positif, negatif, atau netral terhadap penggunaan motor listrik pada ojek online. Data akan dibagi menjadi data train dan data test dengan perbandingan 80:20.

6. Evaluasi Metode SVM

Mengukur performa metode dalam bentuk *accuracy*, *precision*, *recall*, dan *F1-score*. *Confusion matrix* digunakan untuk mengukur sejauh mana model dapat mengklasifikasikan komentar dengan benar.

- a. *Accuracy*, menunjukkan persentase klasifikasi yang benar.
- b. *Precision*, mengukur seberapa banyak jumlah data yang benar positif dari semua hasil prediksi yang positif.
- c. *Recall*, mengukur seberapa banyak data yang prediksi benar positif dari hasil seluruh benar positif.
- d. *F1-score*, memberi gambaran tentang performa model, terutama ketika ada ketidakseimbangan antara kelas positif dan negatif.

7. Menentukan Jumlah Topik LDA

Setelah model SVM mengklasifikasikan komentar ke dalam kategori sentimen (positif, negatif, netral), setiap sentimen tersebut dianalisis secara terpisah menggunakan LDA dengan *coherence score* untuk menentukan jumlah topik optimal dalam komentar. Semakin tinggi *coherence score* maka semakin baik kualitas topik yang terbentuk. Proses ini dilakukan dengan mencoba berbagai jumlah topik dan memilih jumlah yang menghasilkan *coherence score* paling tinggi.

8. Pemodelan Topik

Menyajikan hasil pemodelan topik untuk melihat topik apa saja yang sering muncul dalam komentar positif, negatif, maupun netral. Pemodelan topik ini dapat mengidentifikasi isu-isu utama yang dibahas dalam komentar, seperti aspek yang dianggap bermanfaat dalam kategori sentimen positif, serta keluhan apa yang sering muncul dalam sentimen negatif. Hasil pemodelan topik ini disajikan dalam bentuk tabel untuk memberi gambaran yang jelas mengenai tema yang sering muncul dalam komentar.

9. Analisis Hasil

Setelah diperoleh topik dengan kategori positif, negatif, dan netral, dilakukan analisis terhadap masing-masing topik dalam setiap kategori sentimen. Analisis ini dengan menghitung jumlah komentar yang masuk ke

dalam topik dengan kategori sentimen positif, negatif, dan netral. Mengidentifikasi topik dengan jumlah komentar sentimen positif, negatif, dan netral tertinggi untuk mengetahui topik yang sering dibahas dalam masing-masing kategori sentimen. Hasil analisis kemudian dibuat visualisasi menggunakan diagram untuk menunjukkan distribusi sentimen. Dengan analisis ini didapat kesimpulan mengenai persepsi masyarakat terhadap penggunaan motor listrik untuk ojek online, sehingga dapat digunakan untuk perbaikan dan peningkatan kualitas layanan dalam adopsi motor listrik dalam sektor transportasi online.

PERPUSTAKAAN
UNIVERSITAS JENDERAL ACHMAD YANU
YOGYAKARTA