

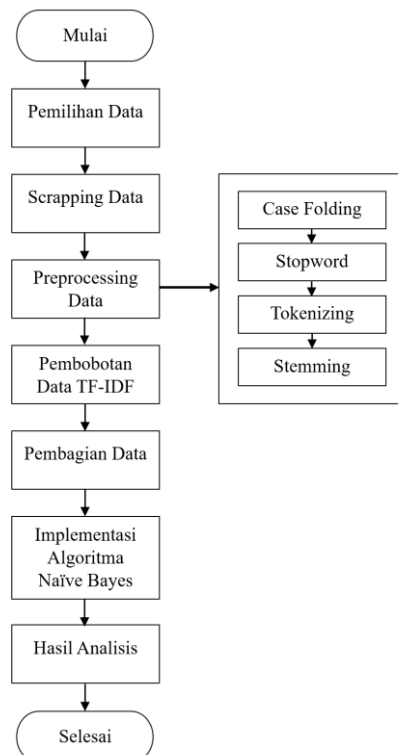
BAB III METODE PENELITIAN

3.1 Objek Penelitian

Objek penelitian ini adalah ulasan pengguna aplikasi *e-commerce* Lazada dan Tokopedia yang diambil dari *Google Play Store*. Penelitian fokus pada analisis sentimen menggunakan Algoritma *Naive Bayes* untuk melakukan pengolahan data terhadap aplikasi tersebut. Data dikumpulkan melalui metode *web scraping* dan diolah menggunakan Google Colaboratory. Penelitian dilakukan secara daring dalam rentang waktu data Februari 2024 hingga Januari 2025, mencakup tahap pengumpulan, pengolahan, hingga pemisahan data, dengan lingkup penelitian dibatasi pada ulasan di *Google Play Store*.

3.2 Prosedur Penelitian

Prosedur penelitian ini dirancang secara terstruktur untuk menjamin keakuratan data dan menyediakan langkah-langkah yang dapat direplikasi oleh peneliti lain di masa mendatang. Diagram alur penelitian dapat dilihat pada Gambar 3.1, dengan setiap tahapan sebagai berikut:



Gambar 3.1 Diagram Alur Penelitian

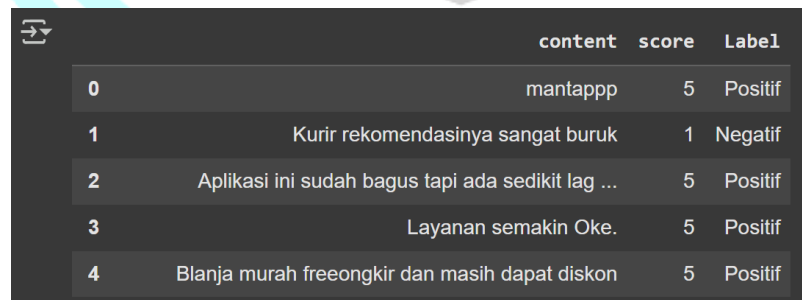
Gambar di atas menampilkan diagram alur penelitian saya untuk menganalisis sentimen pada ulasan pengguna aplikasi Lazada dan Tokopedia menggunakan Algoritma *Naive Bayes*. Tahapan dalam diagram alur penelitian ini meliputi:

1. Pemilihan Data

Pada tahap ini, data ulasan pengguna aplikasi *e-commerce* Lazada dan Tokopedia dipilih dari *platform Google Play Store*. Data yang diambil meliputi teks ulasan, peringkat, dan tanggal unggahan. Pemilihan data dilakukan dengan memastikan bahwa ulasan yang diambil relevan dan sesuai dengan tujuan penelitian, yakni untuk analisis sentimen, serta mencakup jumlah data yang memadai untuk proses selanjutnya.

2. Scrapping Data

Web scraping adalah proses otomatisasi pengambilan data dari halaman web. Dalam konteks ini, Sebanyak 2000 ulasan dikumpulkan melalui metode *web scraping* di *Google Play Store* dengan alat *Google Colaboratory*. Data tersebut kemudian melalui proses *preprocessing data* tersebut kemudian siap untuk dianalisis lebih mendalam, di mana algoritma *Naive Bayes* diterapkan untuk klasifikasi sentiment.



	content	score	Label
0	mantapp	5	Positif
1	Kurir rekomendasinya sangat buruk	1	Negatif
2	Aplikasi ini sudah bagus tapi ada sedikit lag ...	5	Positif
3	Layanan semakin Oke.	5	Positif
4	Blanja murah freeongkir dan masih dapat diskon	5	Positif

Gambar 3.2 Dataset

3. Preprocessing Data

Setelah data ulasan aplikasi Lazada berhasil dikumpulkan melalui *web scraping*, langkah selanjutnya adalah melakukan *text preprocessing*. *Text preprocessing* membersihkan dan mempersiapkan data teks agar siap untuk dianalisis (Hananto, 2024). Setelah itu, sentiment analisis akan dilakukan untuk mengekstrak sentimen setiap ulasan pengguna. Kegiatan ini mencakup sebagai berikut :

- a. *Case Folding* : *Folding case* mengubah semua huruf menjadi huruf kecil.
- b. *Stopword* : Kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh *stopword* dalam bahasa Indonesia adalah “yang”, “dan”, “di”, “dari”, dll. Makna di balik penggunaan *stopword* yaitu dengan menghapus kata-kata yang memiliki informasi rendah dari sebuah teks, kita dapat fokus pada kata-kata penting sebagai gantinya.
- c. *Tokenization* : Teks dihapus dari tanda baca dan dibagi menjadi bagian yang lebih kecil berdasarkan spasi karakter selama proses tokenisasi.
- d. *Stemming* : Proses *stemming* menemukan kata dasar dan menghilangkan semua imbuhan. Tujuan *stemming* adalah untuk menghilangkan kata-kata dengan pengejaan yang buruk. Algoritma *stemming* yang digunakan dalam setiap bahasa berbeda. Selain itu, *stemming* digunakan untuk meningkatkan keseragaman data dan menurunkan jumlah kata yang ada.

text_clean	text_StopWord	text_tokens	text_steamindo
mantapp	mantapp	['mantapp']	mantapp
kurir rekomendasinya sangat buruk	kurir rekomendasinya buruk	['kurir', 'rekomendasinya', 'buruk']	kurir rekomendasi buruk
aplikasi ini sudah bagus tapi ada sedikit lag ...	aplikasi bagus lag aplikasi mengganggu tolong d...	['aplikasi', 'bagus', 'lag', 'aplikasi', 'meng...']	aplikasi bagus lag aplikasi mengganggu tolong d...
layanan semakin oke	layanan oke	['layanan', 'oke']	layan oke
blanja murah freeongkir dan masih dapat diskon	blanja murah freeongkir diskon	['blanja', 'murah', 'freeongkir', 'diskon']	blanja murah freeongkir diskon

Gambar 3.3 Hasil *Preprocessing*

4. Pembobotan Data TF-IDF

Pembobotan data TF-IDF (*Term Frekuensi-Inverse Document Frekuensi*), yang mengukur pentingnya sebuah kata dalam sebuah dokumen berdasarkan frekuensinya. *Term Frekuensi* (TF) menghitung seberapa sering sebuah kata muncul dalam dokumen tertentu, sedangkan *Inverse Document Frekuensi* (IDF) memberikan bobot yang lebih rendah pada kata-kata yang sering muncul di banyak dokumen, sehingga lebih menonjolkan kata-kata yang spesifik dan relevan dengan konteks dokumen tersebut.

5. Pembagian Data

Pada tahap pembagian data ini dilakukan pembagian data yang digunakan dalam penelitian ini untuk memisahkan *dataset* menjadi dua

bagian utama, yaitu data latih (data pelatihan) dan data uji (data pengujian). *Hold-out* digunakan untuk membagi data menjadi data uji dan data latih. Data latih digunakan untuk membangun dan melatih model Algoritma *Naive Bayes*, sedangkan data uji digunakan untuk memulai performa model dalam mengklasifikasikan sentimen pada ulasan pengguna. Teknik pemisahan data ini memastikan bahwa model yang dihasilkan berfungsi baik.

6. *Multinomial Naive Bayes*

Multinomial Naive Bayes adalah salah satu varian dari algoritma *Naive Bayes* yang digunakan untuk klasifikasi data teks. Ini menggunakan probabilitas Bayes, yang menentukan kemungkinan bahwa data akan dimasukkan dalam kategori tertentu berdasarkan distribusi kata dalam dokumen. Metode ini mengasumsikan bahwa kata-kata dalam dokumen bersifat independen, sehingga memungkinkan analisis dengan efisien. Selain itu, *Multinomial Naive Bayes* mengklasifikasikan dokumen secara terpisah tanpa menghubungkan hubungan yang ada di antara dokumen, menghasilkan hasil yang lengkap dari analisis dokumen. Berikut ini adalah persamaan yang digunakan dalam algoritma ini. Memakai perintah *python* yaitu :

```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix

clf = MultinomialNB()
clf.fit(X_train, y_train)
predicted = clf.predict(X_test)
```

Gambar 3.4 Kode Perintah

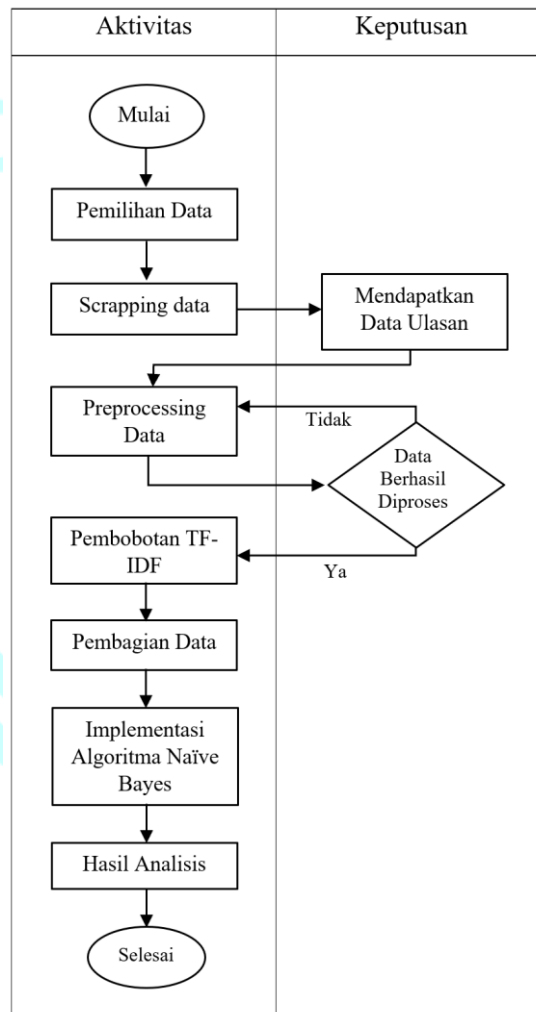
7. Hasil Analisis

Hasil klasifikasi sentimen dari ulasan yang telah melalui proses analisis akan disajikan. Data yang diperoleh dari hasil klasifikasi ini dimanfaatkan untuk menganalisis ulasan pengguna aplikasi Lazada dan Tokopedia di *Play Store*. Analisis ini bertujuan untuk memahami pandangan umum pengguna terhadap kedua aplikasi tersebut berdasarkan evaluasi sentimen yang dilakukan. Dengan demikian, hasil dari klasifikasi sentimen ini dapat

memberikan informasi penting bagi pengembang untuk meningkatkan kualitas layanan dan merespons kebutuhan pengguna dengan lebih baik.

3.3 Activity Diagram

Berikut disajikan *Activity Diagram* yang menggambarkan alur proses penelitian secara sistematis untuk mencapai tujuan yang telah ditetapkan.



Gambar 3.5 Aktivitas Penelitian

Berdasarkan gambar di atas, alur penelitian digambarkan melalui diagram aktivitas yang menjelaskan tahapan-tahapan sistematis dalam implementasi Algoritma *Naïve Bayes* untuk analisis sentimen ulasan pengguna aplikasi Lazada dan Tokopedia di *Google Play Store*. Proses dimulai dengan tahap pemilihan data, di mana ulasan pengguna yang relevan dipilih berdasarkan kebutuhan penelitian untuk memastikan kelengkapan dan kecocokan data. Setelah itu, data yang telah

terpilih dikumpulkan melalui metode *web scraping*. Mengambil data berupa teks ulasan, peringkat, dan tanggal unggahan dari *platform* tersebut. Setelah data berhasil didapatkan, tahap berikutnya adalah *preprocessing data*, yang bertujuan untuk membersihkan dan mempersiapkan data agar siap digunakan dalam analisis. Proses ini mencakup penghapusan duplikasi data, penghilangan karakter khusus atau simbol yang tidak relevan, tokenisasi, normalisasi teks, serta penghapusan *stopwords*. Selanjutnya, data diproses menggunakan pembobotan TF-IDF, teknik ini memungkinkan pengukuran sebuah kata dalam ulasan tertentu dengan mempertimbangkan distribusi kata tersebut di seluruh *dataset*.

Berikutnya adalah pembagian data, di mana data dibagi menjadi dua bagian: data latih dan uji. Metode *Hold-Out* digunakan untuk memastikan bahwa data latih digunakan dalam membangun model, sementara data uji digunakan untuk mengevaluasi performa model yang telah dilatih.

Pada tahap implementasi, Algoritma *Naive Bayes* diterapkan pada data latih untuk membangun model klasifikasi sentimen. Model ini kemudian diuji dengan data uji untuk mengukur akurasi, presisi, *recall*, dan *f1-score*, sehingga dapat dinilai seberapa baik algoritma ini dalam menganalisis sentimen ulasan pengguna. Hasil dari seluruh tahapan tersebut akan diinterpretasikan untuk memberikan wawasan mengenai pola sentimen pengguna terhadap aplikasi Lazada dan Tokopedia. Alur proses ini dirancang untuk memastikan bahwa setiap langkah berjalan secara terstruktur, menjelaskan secara rinci keseluruhan alur yang dilakukan, serta memberikan gambaran yang jelas mengenai tahapan-tahapan yang ditempuh dalam penelitian.