BAB 3
METODE PENELITIAN

Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi Maxim dengan menerapkan metode Naive Bayes. Data yang dianalisis dalam penelitian ini berasal dari ulasan pengguna yang diambil dari Google Play Store. Proses penelitian diawali dengan mengidentifikasi permasalahan yang terungkap melalui analisis ulasan pengguna. Kemudian, dilakukan tahap preprocessing data untuk membersihkan dan mempersiapkan data sebelum dilakukan analisis sentimen. Penelitian ini diharapkan dapat menyajikan informasi yang bermanfaat bagi para pengguna.

3.1 BAHAN DAN ALAT PENELITIAN

Penelitian yang sedang dikembangkan adalah analisis sentimen pengguna terhadap aplikasi Maxim dengan metode Naive Bayes. Data yang dianalisis berasal dari komentar pengguna aplikasi maxim diperoleh dari Google Play Store. Awal mula untuk mengerjakan penelitian ini dengan mengindentifikasikan masalah yang muncul melalui analisis ulasan pengguna. Selanjutnya, dilakukan tahap preprocessing data untuk membersihkan dan mempersiapkan data sebelum dilaksanakan analisis sentimen.

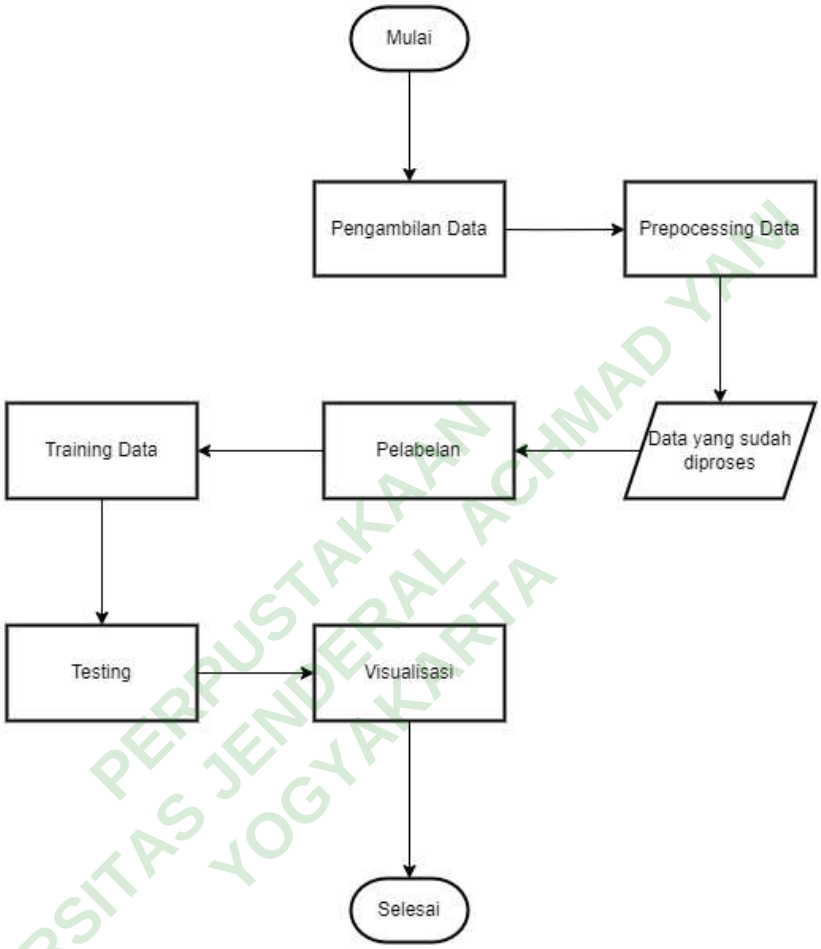
Untuk menjalankan penelitian ini dibutuhkan sistem dan perangkat lunak pada komputer dengan spesifikasi yang memadai, serta dilengkapi dengan akses jaringan Internet. Adapun spesifikasi dan aplikasi yang digunakan antara lain :

1. Sistem operasi : Windows 11
2. Python 3.11
3. Google Play Store
4. Google Colab
5. Googleplay scraper 1.2.7

PERPUSTAKAAN UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA

3.2 JALAN PENELITIAN

Tahapan-tahapan yang ditempuh dalam menganalisis sentimen ulasan pengguna pada aplikasi Maxim dapat dilihat pada Gambar 3.1.



Gambar 3.1 Jalan Penelitian

3.2.1 Pengambilan Data

Pada tahap pengambilan data, dilakukan pengumpulan ulasan aplikasi Maxim dari Google Play Store. Proses ini dilaksanakan menggunakan Google Colab dengan memanfaatkan Google Play Scraper sebagai alat untuk mengumpulkan data secara otomatis dari Google Play Store. Data ulasan yang telah didapatkan akan disimpan dalam format file *Comma Separated Values* sehingga dapat diakses dan dianalisis menggunakan perangkat lunak seperti Microsoft Excel. Berikut kode yang digunakan :

PERPUSTAKAAN UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA

```
from google_play_scraper import Sort, reviews
result, continuation_token = reviews(
    'com.taxsee.taxsee',
    country='id',
    count=200000,
 filter_score_with=None
```

Hasil data yang telah diperoleh dari hasil *scraping* dapat dilihat pada Gambar 3.2.

	reviewId	userName	userImage	content	score
0	7a61bcdc-0f24-4727-be0d-404948a996c9	Pengguna Google	https://play-lh.googleusercontent.com/EGemol2N...	Plusnya lebih murah ketimbang ojol lainnya. Mi...	1
1	4faf3670-0581-44c4-aa7d-27377b0f1fb1	Pengguna Google	https://play-lh.googleusercontent.com/EGemol2N...	Aplikasi sangat membantu. Tapi tolong diperbai...	3
2	8d3f1fad-4d84-41c5-9e1e-feab5acb3f75	Pengguna Google	https://play-lh.googleusercontent.com/EGemol2N...	Aplikasi ini bagus sebenernya, harga2 nya pada...	5
3	c4728306-409b-44bd-b1ad-0da54363b484	Pengguna Google	https://play-lh.googleusercontent.com/EGemol2N...	Saya baru pertama kali coba top up saldo maxim...	1
4	6fe11b23-53a6-410f-8e12-af2d2965b2e3	Pengguna Google	https://play-lh.googleusercontent.com/EGemol2N...	Saya suka dengan aplikasi ini, namun setiap sa...	3
...	...	...	...	...	...
199995	76a9da65-faa5-4fdd-a450-da207dca947b	Eden Fandu	https://play-lh.googleusercontent.com/a-/ALV-U...	Jelek	1
199996	31d4d050-e2f2-4606-ba03-ba176413b390	Agus Riadi	https://play-lh.googleusercontent.com/a/ACg8oc...	mantaf	1
199997	ba4772da-23d8-4ccd-a228-b8a6a7625cd0	Ihwan Angrh	https://play-lh.googleusercontent.com/a/ACg8oc...	KARETTTTTTT	1
199998	4b24d686-2aea-4f94-bbd2-073b35db02db	Arnold Ayomi	https://play-lh.googleusercontent.com/a-/ALV-U...	bagus	1
199999	8ba673af-d030-4862-9299-332366135753	roslina gulo	https://play-lh.googleusercontent.com/a/ACg8oc...	Bagus 😍👍👍	5

200000 rows × 11 columns

Gambar 3.2 Hasil Scraping

3.2.2 Tahap Prepocessing Data

Pada tahap pra-pemrosesan, data ulasan pengguna Maxim yang didapatkan berupa teks yang terdiri dari berbagai elemen seperti huruf, angka, tanda baca, emotikon, dan lain sebagainya. Mengingat bahwa data teks adalah data yang tidak terstruktur, maka diperlukan langkah-langkah pra-pemrosesan terlebih dahulu sebelum data tersebut dapat diolah lebih lanjut. Pra-pemrosesan data bertujuan untuk menjadikan dataset yang awalnya berupa teks tidak terstruktur menjadi data yang terstruktur. Beberapa proses yang dilalui dalam tahapan pra-pemrosesan data antara lain sebagai berikut.

1. *Data Cleaning*

   Data cleaning adalah proses untuk menghapus karakter selain huruf, seperti angka dan tanda baca. Dapat dilihat pada cuplikan kode berikut:

```
my_df[new_text_field_name] = my_df[new_text_field_name]
.apply(lambda elem: re.sub(r"(@[A-Za-z0-9]+)
|([^0-9A-Za-z\t])|(\w+:\/\/\S+)|^rt|http.+?", "", elem))
```

   Proses *data cleaning* menghasilkan data seperti pada Tabel 3.1

Tabel 3.1 *Data Cleaning*

Data Mentah	**Hasil Data Cleaning**
Sangat baik dan ramah, Terimakasih Pak	Sangat baik dan ramah Terimakasih Pak
Sangatlah bagus mksih 🙏🙏❤️	Sangatlah bagus mksih
Pelayanannya sangat baik.. Recommended bgt	Pelayanannya sangat baik Recommended bgt
Bapak ny baik ,byk" rezeki ny bpk byk yg order...	Bapak ny baik byk rezeki ny bpk byk yg order

2. *Case Folding*

   Case Folding merupakan salah satu teknik dalam pra-pemrosesan data dengan tujuan agar merubah keseluruhan huruf dalam data menjadi huruf kecil. Hal ini penting untuk memastikan konsistensi dalam analisis teks, terutama dalam menghindari perbedaan yang disebabkan oleh variasi huruf besar dan kecil. Berikut cuplikan kodenya kodenya:

UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA PERPUSTAKAAN

```
my_df[new_text_field_name] = my_df[text_field].str.lower()
```

Proses *Case Folding* menghasilkan data yang ditunjukkan pada Tabel 3.2.

Tabel 3.2 *Case Folding*

Data Cleaning	**Hasil Case Folding**
Sangat baik dan ramah terimakasih pak	sangat baik dan ramah terimakasih pak
Pelayanannya sangat baik Recommended bgt	pelayanannya sangat baik recommended bgt
Pelayanannya Recommended bgt	pelayanannya recommended bgt
Bapak ny baik byk rezeki ny bpk byk yg order	bapak ny baik byk rezeki ny bpk byk yg order

3. *Stopword Removal*.

   Stopword Removal adalah salah satu tahapan untuk menghilangkan kata-kata yang tidak relevan dalam kalimat berdasarkan daftar *stopword*. Berikut adalah contoh kode yang dapat digunakan untuk melakukan *Stopword Removal*:

```
nltk.download('stopwords')
stop = stopwords.words('indonesian')
data_clean['text_StopWord'] = data_clean['text_clean'].
   apply(lambda x:' '.join([word for word in x.split()
      if word not in (stop)]))
```

Proses *Stopword Removal* menghasilkan data yang ditunjukkan pada Tabel 3.3

Tabel 3.3 *Stopword Removal*

Case Folding	**Hasil Stopword Removal**
sangat baik dan ramah terimakasih pak	ramah terimakasih
sangatlah bagus mksih	bagus mksih
pelayanannya sangat baik recommended bgt	pelayanannya recommended bgt
bapak ny baik byk rezeki ny bpk byk yg order	ny byk rezeki ny bpk byk yg order
sangatlah bagus mksih	bagus mksih

UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA PERPUSTAKAAN

4. *Tokenizing*.

   Tokenizing adalah proses dalam pemrosesan teks yang panjang kemudian dipecah menjadi kata-kata atau unit-unit lain yang lebih kecil, yang dapat dianalisis secara individu. Berikut kode yang digunakan:

```
from nltk.tokenize import sent_tokenize, word_tokenize
data_clean['text_tokens'] data_clean['text_StopWord'].
       apply(lambda x: word_tokenize(x))
data_clean.head()
```

   Proses *Tokenizing* menghasilkan data yang ditunjukkan pada Tabel 3.4

Tabel 3.4 *Tokenizing*

Stopword Removal	**Hasil Tokenizing**
ramah terimakasih	[ramah, terimakasih]
bagus mksih	[bagus, mksih]
pelayanannya recommended bgt	[pelayanannya, recommended, bgt]
ny byk rezeki ny bpk byk yg order aamiin	[ny, byk, rezeki, ny, bpk, byk, yg, order, aamiin]

5. *Stemming*.

   Stemming adalah proses pengubahan bentuk kata kedalam bentuk dasarnya. Kata-kata yang telah melewati proses *stopword removal* kemudian akan diubah ke dalam bentuk dasar dengan menghilangkan imbuhan depan maupun imbuhan belakang yang ada. Berikut kode yang digunkan:

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
def stemmed_wrapper(term):
return stemmer.stem(term)
       for document in data_clean['text_tokens']:
           for term in document:
               if term not in term_dict:
            term_dict[term] = ' '
```

   Proses *Stemming* menghasilkan data yang ditunjukkan pada Tabel 3.5.

UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA PERPUSTAKAAN

Tabel 3.5 *Stemming*

Tokenizing	Hasil Stemming
[ramah, terimakasih]	ramah terimakasih
[bagus, mksih]	bagus mksih
[pelayanannya, recommended, bgt]	layan recommended bgt
[ny, byk, rezeki, ny, bpk, byk, yg, order, aamiin]	ny byk rezeki ny bpk byk yg order aamiin

3.2.3 Tahap Pelabelan

Pada tahap pelabelan, setiap kata dalam dokumen diberi label sentimen positif atau negatif untuk mengidentifikasi dan mengklasifikasikan polaritas sentimen dalam teks. Proses ini melibatkan peninjauan manual berdasarkan aturan yang telah ditentukan untuk memastikan bahwa setiap ulasan diberi label dengan tepat, yang nantinya akan digunakan sebagai dasar dalam pelatihan model. Pada tahap pelabelan ini menggunakan model Bidirectional Encoder Representations from Transformers (BERT) sebagai model untuk melakukan pelabelan secara otomatis. Bidirectional Encoder Representations from Transformers (BERT) adalah model bahasa yang diperkenalkan oleh Google pada tahun 2018. BERT bekerja dalam dua tahap: pre-training dan fine-tuning. Penggunaan model BERT menunjukkan peningkatan akurasi yang signifikan dibandingkan dengan metode tradisional dan menyoroti pentingnya pendekatan dua arah yang digunakan oleh BERT (Halim & Liliana, 2022). Berikut kode yang digunakan untuk melakukan pelabelan dengan BERT:

```
from transformers import pipeline
sentiment_classifier = pipeline(
       'sentiment-analysis', model=
              'indobenchmark/indobert-large-p1')
data['label'] = data['cleaned_review']
       .apply(lambda x: sentiment_classifier(x)[0]
              ['label'].lower())
data = data[data['label'].isin(['positive', 'negative'])]
print(data['label'].value_counts())
```

PERPUSTAKAAN UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA

Hasil pelabelan menggunakan model BERT dari 200.000 data menghasilkan 167.213 data positif dan 32.787 data negatif. Berikut hasil, dapat dilihat pada Tabel 3.6.

Tabel 3.6 Pelabelan

Text_Steming	**Hasil Pelabelan Dengan BERT**
ramah terimakasih	Positive
bagus mksih	Positive
layan recommended bgt	Positive
lama jemput nya	Negative

3.2.4 Tahap Training Data

Pada tahap training data, dataset dibagi menjadi dua bagian: training dataset dan testing dataset. Training dataset digunakan untuk mengajarkan model Naive Bayes Classification. Proses pelatihan melibatkan penggunaan ulasan yang telah diberi label untuk membangun model yang dapat mengkategorikan ulasan baru sebagai positif atau negatif berdasarkan pola yang telah dipelajari dari data pelatihan. Pada tahap ini data training yang digunakan sebanyak 80% dari jumlah keseluruhan data yakni 160000 data. Untuk lebih jelasnya dapat dilihat pada Tabel 3.7.

Tabel 3.7 *Training* Data

Paramater	**Jumlah Data**
X_Test	160000
Y_Test	160000

3.2.5 Tahap Testing

Pada tahap testing, Mengukur akurasi model yang telah dilatih dilakukan sebagai bagian dari evaluasi. Pengujian ini menggunakan confusion matrix untuk menghitung akurasi, presisi, recall, dan F1-score. Untuk mendapatkan akurasi dilakukan pembandingan jumlah prediksi yang benar dengan total prediksi,

PERPUSTAKAAN UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA

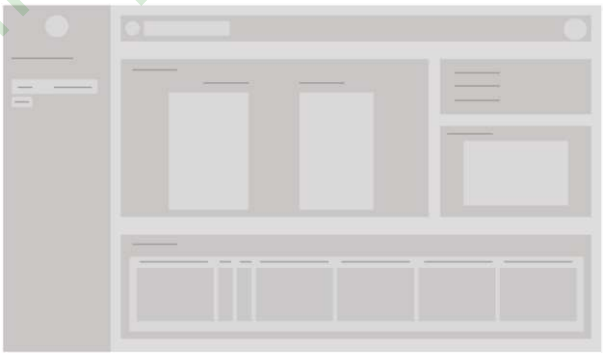
sedangkan metrik lainnya memberikan informasi rinci tentang kesalahan dan keberhasilan prediksi. Dengan demikian, confusion matrix memberikan gambaran komprehensif tentang performa model dan membantu mengidentifikasi serta memperbaiki kelemahan model. Pada tahap ini data testing yang digunakan sebanyak 20% dari jumlah keseluruhan data yakni 40000 data. Dapat ditunjukkan pada Tabel 3.8.

Tabel 3.8 *Testing* Data

Paramater	**Jumlah Data**
X_Test	40000
Y_Test	40000

3.2.6 Rancangan Visualisasi Data

Pada tahap rancangan visualisasi data ini hasil analisis sentimen yang telah dilakukan akan ditampilkan dalam bentuk dashboard interaktif. Dashboard ini akan dibangun menggunakan framework Flask. Dashboard akan menyajikan visualisasi data yang memungkinkan pengguna untuk memahami sentimen ulasan pengguna aplikasi Maxim dengan mudah. Berikut wireframe desain yang dapat dilihat pada Gambar 3.3.



Gambar 3.3 Wireframe

Tampilan wireframe pada Gambar 3.3 digunakan untuk menampilkan visualisasi hasil analisis sentimen negatif dan positif pengguna maxim pada pada Google Play Store. Serta menampilkan hasil evaluasi berupa akurasi, presisi, *recall*, dan *F1 Score.*

PERPUSTAKAAN UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA