

## BAB 3 METODE PENELITIAN

### 3.1 Metode Penelitian

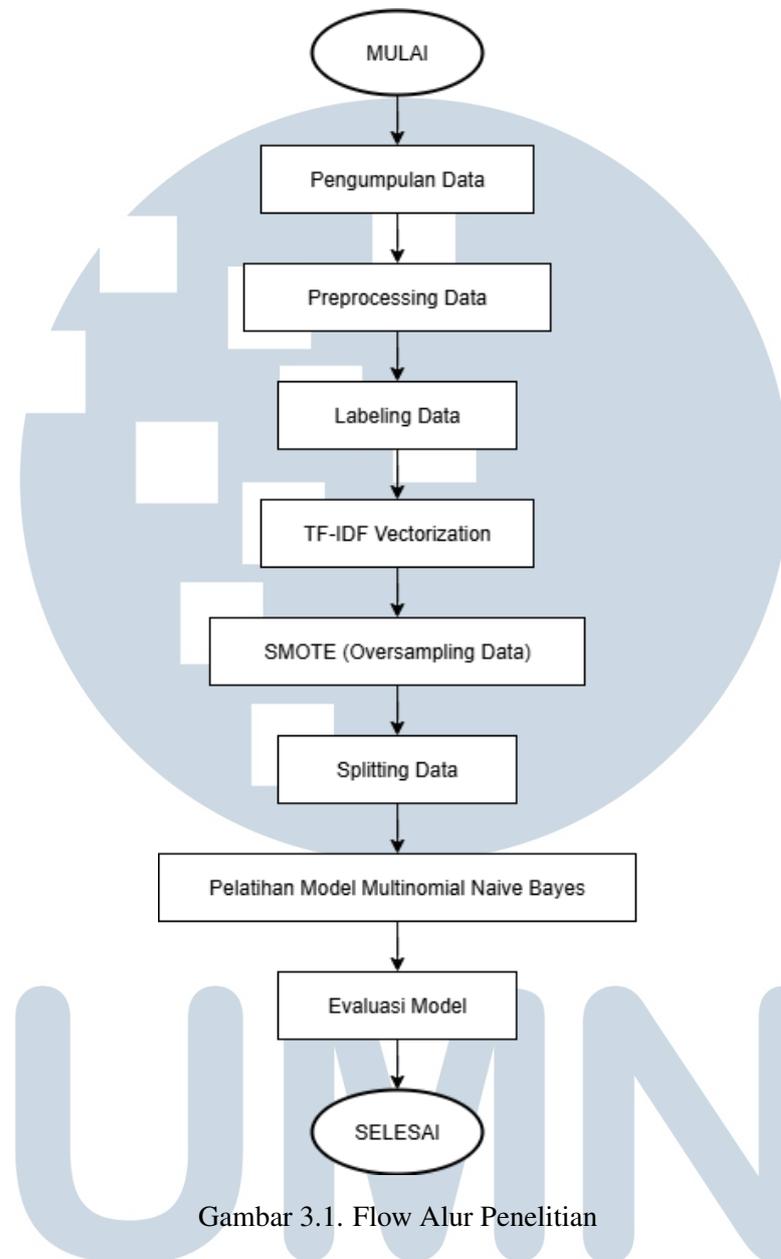
Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi berbasis pembelajaran mesin. Algoritma yang digunakan adalah *Naïve Bayes*, yang cocok untuk klasifikasi teks, terutama dalam analisis sentimen. Penelitian ini bertujuan untuk mengetahui opini publik terhadap pelatih baru Tim Nasional Sepak Bola Indonesia berdasarkan komentar di *platform* Instagram.

Langkah-langkah yang dilakukan meliputi pengumpulan data, *preprocessing* teks, pelabelan sentimen menggunakan *TextBlob*, vektorisasi data dengan TF-IDF, penyeimbangan data menggunakan SMOTE, klasifikasi menggunakan algoritma *Multinomial Naïve Bayes*, dan evaluasi model menggunakan *confusion matrix* dan *classification report*.

### 3.2 Alur Penelitian

Berikut adalah alur yang dilakukan dalam penelitian ini pada Gambar 3.1 .





Gambar 3.1. Flow Alur Penelitian

### 3.3 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari media sosial Instagram. Komentar-komentar dikumpulkan dari beberapa unggahan akun resmi Tim Nasional Sepak Bola Indonesia (@timnasindonesia dan @pssi) dan akun media umum Folkative (@folkative) yang membahas pertandingan Indonesia melawan Australia dan Bahrain dalam rangka Kualifikasi Piala Dunia 2026. Berikut adalah proses pengumpulan data yang dilakukan secara manual.

1. Menyalin komentar dari Instagram  
Komentar yang secara eksplisit menyampaikan opini publik terhadap pelatih baru Timnas Indonesia disalin secara manual dari kolom komentar unggahan terkait.
2. Menempelkan komentar ke aplikasi Notepad  
Komentar terlebih dahulu ditempel ke Notepad agar menghilangkan karakter khusus, bentuk, atau simbol aneh yang mungkin muncul saat disalin langsung dari Instagram, menghasilkan teks yang bersih.
3. Menyalin komentar ke Excel  
Setelah dibersihkan, komentar disalin kembali ke aplikasi Microsoft Excel dan disusun ke dalam bentuk tabular dengan kolom 'id' dan 'komentar'.
4. Menyimpan file ke format .CSV  
Dataset disimpan dalam format .csv agar kompatibel dengan pustaka Python yang digunakan dalam proses analisis.

Proses ini menghasilkan dataset akhir berisi komentar-komentar mentah yang siap digunakan dalam tahap *preprocessing*, pelabelan sentimen, dan klasifikasi akhir. Total komentar yang berhasil dikumpulkan dalam penelitian ini adalah sebanyak 1.857 komentar. Berikut urutan alur yang dilalui dalam proses pengumpulan data pada Gambar 3.2 berikut.



Gambar 3.2. Alur Pengumpulan Komentar di Instagram

Penelitian ini tidak menggunakan data primer seperti wawancara atau survei, melainkan murni menggunakan data sekunder dari media sosial. Pengambilan data dilakukan tanpa bantuan API, melainkan dilakukan secara manual dan sistematis sesuai dengan kriteria pemilihan data yang relevan dengan topik penelitian.

### 3.4 Spesifikasi Sistem

Penelitian ini dilakukan dengan menggunakan perangkat keras (*hardware*) dan perangkat lunak (*software*) sebagai berikut.

### 3.4.1 Perangkat Keras

1. Prosesor: Intel Core i5 9th Gen
2. RAM: 8 GB
3. Penyimpanan: SSD 256 GB

### 3.4.2 Perangkat Lunak

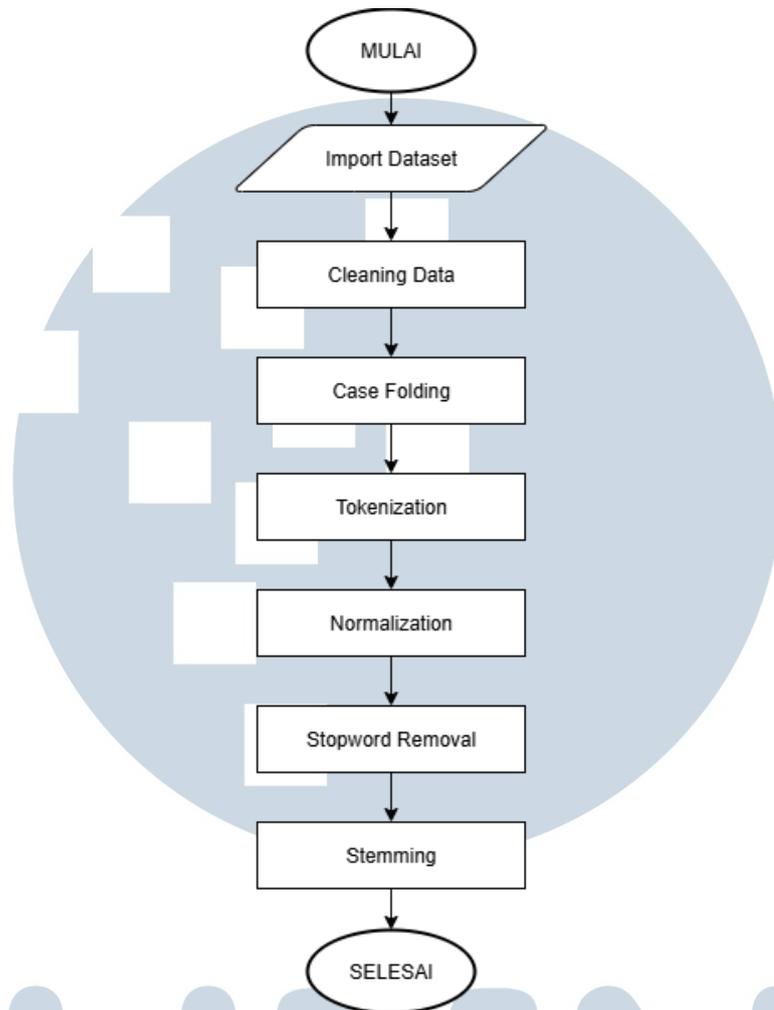
1. Sistem Operasi: Windows 10 64-bit
2. Bahasa Pemrograman: Python 3.11.11
3. Tools: Jupyter Notebook
4. Platform: Anaconda 3
5. Media: Google Chrome

## 3.5 Perancangan Sistem

Perancangan sistem dalam penelitian ini terdiri atas beberapa tahapan utama yang dilakukan secara berurutan untuk menghasilkan model klasifikasi sentimen. Adapun tahapan-tahapan tersebut dijelaskan sebagai berikut.

### 3.5.1 Preprocessing Data

*Preprocessing* merupakan proses awal yang dilakukan untuk membersihkan dan mempersiapkan data teks sebelum digunakan dalam pemodelan. Hasil akhir dari proses *preprocessing* adalah teks yang telah dibersihkan dan disiapkan untuk dianalisis, disimpan dalam kolom '*preprocessed*'. Berikut adalah tahapan *preprocessing* yang dilakukan sesuai dengan yang tertera pada Gambar 3.3.



Gambar 3.3. Flow *Preprocessing* Data

1. *Cleaning*, merupakan proses menghapus karakter yang tidak relevan seperti URL, simbol, angka, dan karakter non-huruf.
2. *Case Folding*, merupakan proses mengubah seluruh huruf menjadi huruf kecil (*lowercase*).
3. *Tokenization*, merupakan proses memecah kalimat menjadi potongan kata.
4. *Normalization*, merupakan proses mengganti kata tidak baku (*slang*) dengan padanan baku jika tersedia.
5. *Stopword Removal*, merupakan proses menghapus kata-kata umum yang tidak memiliki kontribusi penting dalam analisis, seperti "dan", "yang", "di", dan sebagainya.

6. *Stemming*, merupakan proses mengubah kata ke bentuk dasarnya menggunakan library 'Sastrawi'.

### 3.5.2 Labeling Data

Pelabelan data merupakan tahap penting dalam penelitian analisis sentimen, karena menentukan kategori sentimen dari setiap komentar yang menjadi objek penelitian. Dalam penelitian ini, pelabelan dilakukan secara otomatis menggunakan metode berbasis *lexicon*, yaitu dengan memanfaatkan pustaka TextBlob. TextBlob adalah library *Python* yang menyediakan berbagai fitur pemrosesan bahasa alami (*Natural Language Processing*), termasuk analisis sentimen.

TextBlob bekerja dengan cara menghitung nilai polaritas (*polarity*) dari setiap komentar. Nilai polaritas ini berkisar antara -1 hingga +1, dengan interpretasi sebagai berikut.

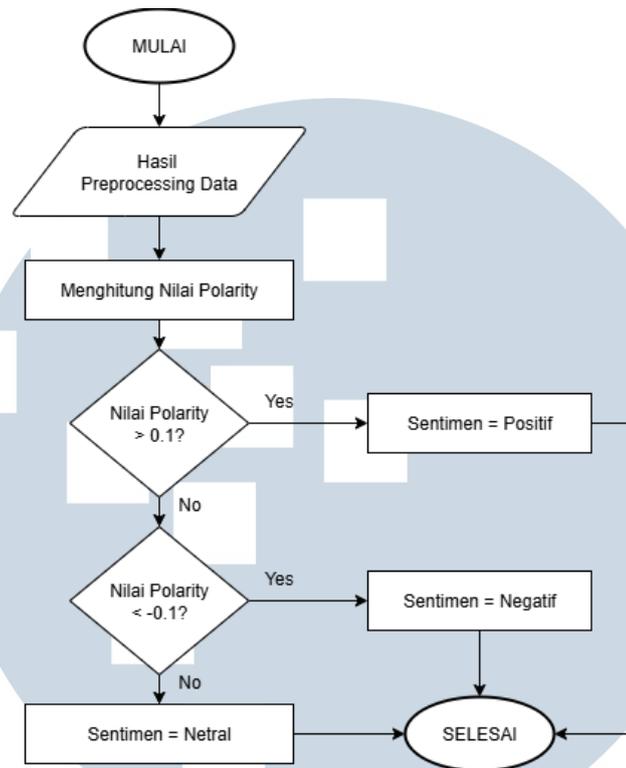
- Nilai mendekati **+1** menunjukkan komentar bernuansa **positif**,
- Nilai mendekati **-1** menunjukkan komentar bernuansa **negatif**,
- Nilai mendekati **0** menunjukkan komentar bersifat **netral**.

Berikut adalah ketentuan klasifikasi sentimen berdasarkan nilai polaritas pada Kode 3.1.

```
1 def get_sentiment(polarity):
2     if polarity > 0.1:
3         return "positif"
4     elif polarity < -0.1:
5         return "negatif"
6     else:
7         return "netral"
```

Kode 3.1: Fungsi pelabelan berdasarkan polaritas TextBlob

Penggunaan ambang batas (*threshold*) sebesar  $\pm 0.1$  bertujuan untuk menghindari ambiguitas pada komentar dengan nilai polaritas yang sangat kecil dan cenderung bersifat netral. Proses pelabelan ini menghasilkan kolom baru dalam dataset yang berisi label sentimen (*positif*, *negatif*, atau *netral*). Label ini kemudian digunakan sebagai target kelas ( $y$ ) pada tahap pelatihan dan pengujian model klasifikasi menggunakan algoritma *Multinomial Naïve Bayes* [27]. Flow dari proses labeling dapat dilihat pada Gambar 3.4.



Gambar 3.4. Flow Labeling Data menggunakan TextBlob

### 3.5.3 TF-IDF Vectorization

Komentar yang telah dipreproses diubah ke bentuk numerik menggunakan teknik TF-IDF. Teknik ini memberikan bobot untuk setiap kata berdasarkan frekuensinya dalam dokumen dan keseluruhan korpus, sehingga menghasilkan representasi fitur yang efektif untuk model klasifikasi teks.

### 3.5.4 SMOTE (Oversampling Data)

Proses ini membantu penelitian karena distribusi label yang tidak seimbang (dominan ke salah satu label sentimen), dilakukan penyeimbangan data menggunakan teknik SMOTE. Teknik ini menghasilkan data sintetis untuk kelas minoritas ('positif' dan 'negatif') berdasarkan distribusi data eksisting, sehingga jumlah data pada ketiga kelas menjadi seimbang.

### 3.5.5 Splitting Data (Train-Test Split)

Data yang telah seimbang dibagi menjadi dua bagian, yaitu Data latih (80%) untuk melatih model klasifikasi dan Data uji (20%) untuk menguji performa model. Pembagian data dilakukan secara stratified agar distribusi label tetap seimbang di kedua bagian.

### 3.5.6 Pelatihan Model Multinomial Naive Bayes

Model klasifikasi yang digunakan adalah *Multinomial Naive Bayes*, yang sangat cocok untuk data hasil vektorisasi teks seperti TF-IDF. Model dilatih menggunakan data latih dan diuji menggunakan data uji yang telah disiapkan.

## 3.6 Evaluasi Model

Evaluasi terhadap model klasifikasi dilakukan untuk mengukur kinerja algoritma dalam mengklasifikasikan komentar ke dalam kategori sentimen yang sesuai. Pengujian dilakukan menggunakan data uji sebesar 20% dari total data hasil balancing.

Hasil evaluasi model ditampilkan dalam bentuk *confusion matrix* dan *classification report* yang mencakup metrik-metrik seperti akurasi, *precision*, *recall*, dan *f1-score*. Berdasarkan hasil pengujian, model *Multinomial Naive Bayes* mampu mengklasifikasikan data dengan tingkat akurasi yang cukup baik dan distribusi prediksi yang seimbang antar kelas sentimen positif, negatif, dan netral.

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A