

## **BAB III**

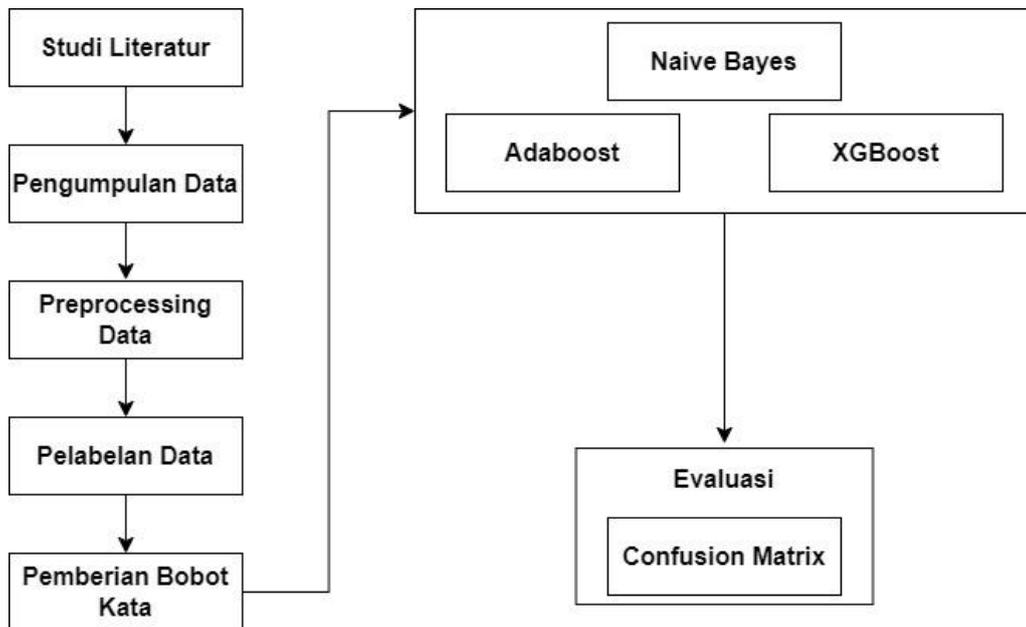
### **METODOLOGI PENELITIAN**

#### **3.1 Jenis dan Sumber Data**

Data yang digunakan dalam penelitian ini adalah data primer. Data primer adalah informasi yang diperoleh langsung dari sumber datanya (Faisal et al., 2020). Data yang digunakan berasal dari media sosial twitter yang diambil dengan metode *web scraping* yang dilakukan menggunakan *library* dari *python* yaitu *snsrape* sebanyak 5.221 data tweet tekstual bahasa indonesia. Berdasarkan 5.221 data tersebut kemudian diurutkan sehingga diperoleh kurun waktu data yang digunakan yaitu 1 Januari 2021 hingga 31 Desember 2022.

#### **3.2 Tahapan Penelitian**

Berikut merupakan tahapan penelitian yang dimulai dari studi literatur, pengumpulan data, data *preprocessing*, pelabelan data, *modelling*, dan *model evaluation* seperti Gambar 3.1



Gambar 3.1 Tahapan Penelitian

### 3.1.1 Studi Literatur

Studi literatur dilakukan dengan mempelajari ilmu dari referensi lain seperti modul, artikel dan jurnal penelitian yang berkaitan dengan masalah penelitian dan untuk mencari dasar teori penelitian sebagai penunjang analisis sentimen yang akan dibuat.

### 3.1.2 Pengumpulan Data

Tahapan selanjutnya dalam penyusunan penelitian ini adalah mengumpulkan data yang relevan terkait penelitian yang akan dilakukan sebagai bahan analisis yang akan digunakan. Pengumpulan data diambil dari tweets yang berhubungan dengan Timnas Futsal Indonesia dari twitter.

### 3.1.3 *Data Preprocessing*

Tahapan selanjutnya adalah tahap *Preprocessing* untuk mempermudah proses analisis data. Pada langkah ini dilakukan melalui beberapa tahapan diantaranya:

1. *Case Folding*

*Case Folding* merupakan tahap penyeragaman dari huruf besar menjadi huruf kecil pada kata, serta menghilangkan seluruh tanda baca pada kalimat.

2. *Tokenization*

*Tokenization* merupakan proses memilah dan memisahkan suatu kalimat menjadi beberapa kata, yang disebut dengan token. Pada proses ini juga berguna untuk mempermudah proses selanjutnya.

3. *Filtering*

*Filtering* adalah proses penghapusan token yang tidak relevan dan tidak memiliki makna pada suatu dataset berdasarkan *database* kamus *stopwords* yang dibentuk.

4. *Stemming*

*Stemming* adalah proses perubahan kata yang berimbuhan menjadi sebuah kata dasar. Tahapan ini dapat mengurangi jumlah kandungan dalam kata.

### 3.1.4 **Pelabelan Data**

Pelabelan data merupakan proses pelabelan data dengan menggunakan library dari *Textblob*. Library ini memudahkan untuk memberikan label secara otomatis dalam jumlah data yang banyak.

### 3.1.5 Pemberian Bobot Kata

Pemberian bobot kata dilakukan untuk menentukan bobot kata pada setiap data ulasan. Nilai bobot suatu kata ditentukan dengan seberapa sering kata muncul dalam setiap ulasan. Metode yang digunakan untuk pembobotan kata yaitu metode ekstraksi fitur TF-IDF.

### 3.1.6 Klasifikasi Data

Klasifikasi data meliputi aktivitas membagi data menjadi dua bagian yaitu data latih dan data uji. Data latih digunakan untuk membangun dan melatih model klasifikasi. Data diuji digunakan untuk menguji model klasifikasi untuk memperoleh tingkat akurasi dari metode klasifikasi. Terdapat tiga kondisi pengujian yang dilakukan pada metode *Naïve Bayes Classifier* yaitu sebelum dilakukan *boosting*, *Naïve Bayes Classifier* menggunakan *Adaboost* dan *Naïve Bayes Classifier* menggunakan *XGBoost*.

### 3.1.7 Evaluasi

Evaluasi meliputi proses menghitung performa dari metode *Naïve Bayes Classifier* menggunakan *confusion matrix*. Berdasarkan hasil evaluasi maka dapat dibandingkan antara sebelum dan sesudah dilakukan optimasi menggunakan *Adaboost* dan *XGBoost*. Selain itu, pada tahap evaluasi juga dilakukan visualisasi untuk melihat kata apa saja yang sering muncul pada setiap sentimennya.