

BAB III

METODOLOGI PENELITIAN

3.1 Objek Penelitian

Dalam penyusunan skripsi ini dilakukan penelitian analisis sentimen di platform media sosial X sebagai sumber informasi utama dalam pengolahan data. Objek penelitian dilakukan terhadap pengguna X dengan kata kunci Skizofrenia.

3.1.1 Instrumen Penelitian

Adapun instrumen penelitian yang digunakan dalam penelitian yaitu:

1. Perangkat Keras

Perangkat keras yang digunakan untuk penelitian ini adalah sebagai berikut: Laptop Acer Aspire A514-55G dengan spesifikasi *processor* 12th Gen Intel(R) Core(TM) i5-1235U 1.30 GHz.

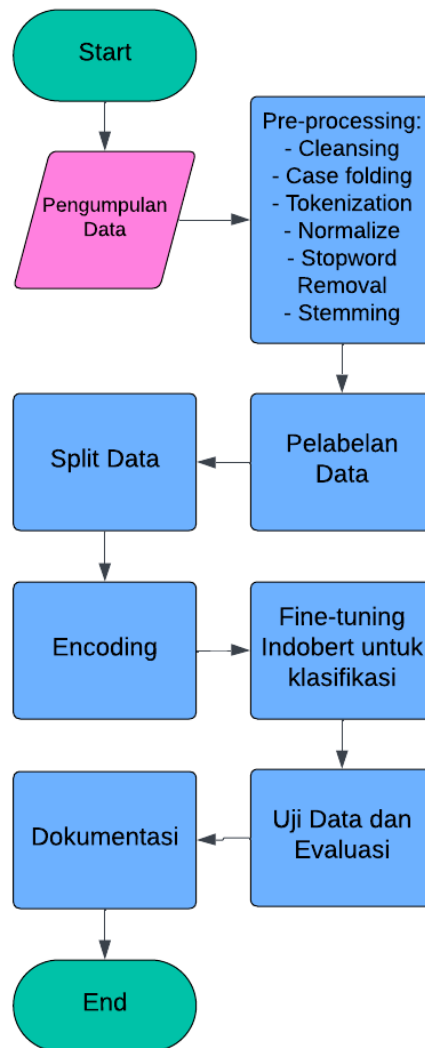
2. Perangkat Lunak

Adapun perangkat lunak yang digunakan pada penelitian ini adalah sebagai berikut:

- a. Windows 11 Home Single Language
- b. Notepad
- c. Google Drive
- d. Google Collab
- e. Python 3.10.12
- f. Social media X/Twitter

3.2 Tahapan Penelitian

Pada sub-bab ini akan dijelaskan proses dan prosedur pelaksanaan langkah-langkah yang dilakukan dalam tugas akhir. Gambaran tahapan penelitian tugas akhir ini disajikan sebagai berikut.



Gambar 3. 1 Tahapan Penelitian

3.2.1 Studi Literatur

Pada tahap studi literatur dilakukan kajian kepustakaan terhadap penelitian terkait analisis sentimen dan *pre-train model Bidirectional Encoder Representations from Transformer* (BERT) yang didapat dari beberapa buku, jurnal internasional, atau penelitian sebelumnya.

3.2.2 Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan skizofrenia selama waktu yang telah ditentukan yang diambil dari media sosial X. Pengambilan

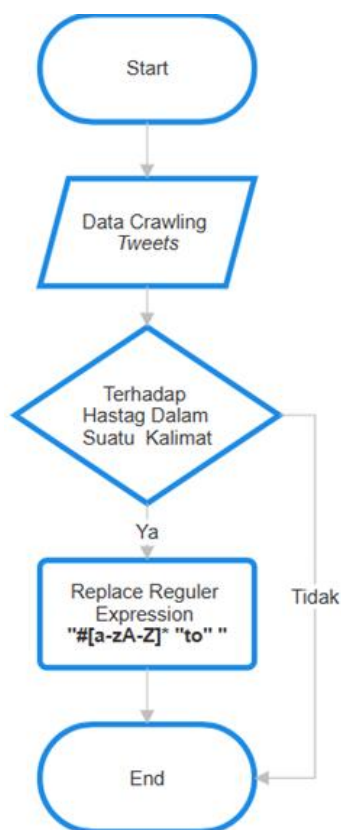
data opini mengenai skizofrenia dilakukan dengan cara melakukan *crawling* data *tweets*. Data opini yang akan dianalisis merupakan *tweets* berbahasa Indonesia.

3.2.3 Pre-processing

Tahap selanjutnya yaitu tahap *pre-processing* data. Tahap ini bertujuan untuk memproses data mentah agar dapat digunakan sebagai data *training* penelitian. Tahapan yang dilakukan pada *pre-processing* data adalah:

a. *Cleansing*

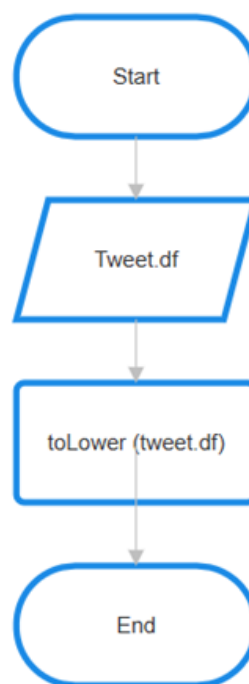
Tujuan dari tahap *cleansing* atau tahap pembersihan data adalah untuk menghilangkan *noise* dan data yang tidak relevan dari dataset, sehingga dapat meningkatkan kualitas model dan performa pemrosesan bahasa alami. Gambar dibawah ini merupakan contoh dari alur proses *cleansing*.



Gambar 3. 2 Diagram Cleansing

b. *Case Folding*

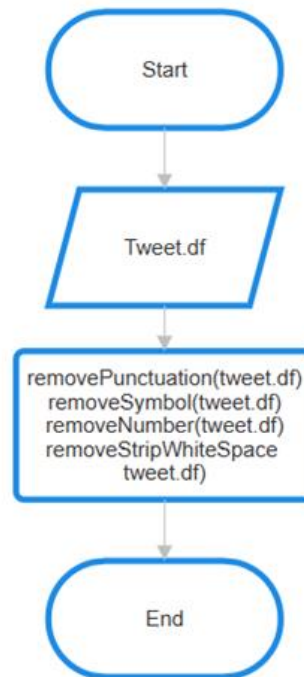
Pada tahap ini, semua huruf diubah huruf kecil. Tahap ini dilakukan karena data yang diperoleh tidak selalu terstruktur dan memiliki kapitalisasi yang konsisten. Oleh karena itu, dilakukan untuk menyamaratakan penggunaan huruf kapital. Gambar dibawah ini merupakan contoh dari alur proses *case folding*.



Gambar 3. 3 Diagram Case Folding

c. *Tokenization*

Tahap *tokenization* pada data *tweet* membagi teks yang panjang menjadi bagian kecil, yang disebut token. Token dapat berupa kata, frasa, atau simbol. Dalam konteks *tweet*, tokenisasi biasanya dilakukan untuk memecah kalimat menjadi kata-kata penyusunnya. Gambar dibawah ini merupakan contoh dari alur proses tokenization.



Gambar 3. 4 Diagram Tokenization

d. *Normalize*

Tahap normalisasi adalah tahap dataset yang memiliki kata-kata baku diubah menjadi kata yang baku atau sesuai dengan ejaan. Ini dilakukan karena banyak kalimat yang menggunakan kata gaul seperti : dmn, tdk, ga, enggak, gak, cpt, srty, dlm, mlm, bgt. Tahap ini bertujuan untuk meningkatkan kualitas dan konsistensi data, sehingga memudahkan model bahasa untuk memproses dan memahami teks dengan lebih baik.

e. *Stopword Removal*

Tahap *stopword removal* data *tweet* menghilangkan kata-kata yang dianggap tidak penting atau tidak menggambarkan isi *tweet*. Pemrosesan ini dilakukan untuk mengurangi noise dan meningkatkan akurasi analisis data *tweet*.

f. *Stemming*

Proses mengurangi kata menjadi bentuk dasarnya. Dalam bahasa Indonesia, *stemming* melibatkan penghapusan awalan, akhiran, dan infiks dari kata.

3.2.4 Pelabelan Data

Data yang dikumpulkan kemudian dilakukan pelabelan menggunakan *lexicon*. Pelabelan data bertujuan agar model dapat belajar sehingga dapat menentukan data opini termasuk dalam kategori sentimen positif, negatif, maupun netral.

3.2.5 Implementasi IndoBERT

Setelah tahap *preprocessing* data, dataset diubah menjadi input yang dapat diterima IndoBERT berupa vektor representasi kata menggunakan tokenizer IndoBERT. Selanjutnya, proses *fine-tuning* dimana *pre-train* model BERT untuk menyesuaikan prioritas model BERT dalam melakukan klasifikasi sentimen.

3.2.6 Uji Coba dan Evaluasi

Setelah melalui tahap implementasi IndoBERT, model diuji dengan menampilkan hasil klasifikasi sentimen dari data pengujian. Klasifikasi sentimen yang dihasilkan dievaluasi menggunakan *confusion matrix*. Hal ini menunjukkan keakuratan dan kinerja IndoBERT dalam melakukan analisis sentimen.

3.2.7 Dokumentasi

Pada tahap akhir, dokumen penelitian dibuat dalam bentuk buku tugas akhir dengan struktur yang telah ditentukan. Saya berharap dengan mendokumentasikan hasil penelitian ini dapat memberikan penjelasan yang mudah dipahami jika penelitian ini dikembangkan lebih lanjut.

3.3 Data Yang Digunakan

Data yang digunakan dalam penelitian ini dikumpulkan dari media sosial X dengan teknik data *crawling*. Metode *crawling* data adalah suatu teknik pengambilan data dari suatu website atau media sosial secara otomatis dengan menggunakan program komputer[19]di simpan ke sebuah *file* berformat .csv yang kemudian diolah ke tahap *preprocessing*.

3.4 Pengujian

Pengujian pada penelitian ini menggunakan metode *confusion matrix* yang menguji performa mesin/model guna memberikan lebih detail klasifikasi yang benar dan salah untuk setiap kategori (data uji dan data prediksi). *Confusion matrix* ini berisi semua informasi mentah tentang prediksi yang dibuat oleh model klasifikasi pada kumpulan data tertentu. Indikator kinerjanya dapat berupa, misalnya *precision*, *recall*, *f1-score*, *support*.

3.4.1 Implementasi Model

Setelah menentukan bahwa model telah dievaluasi secara memadai, model tersebut diimplementasikan pada kumpulan data yang lebih besar. Pada tahap ini, model mulai memberi label pada data yang dikumpulkan sebagai positif, negatif, dan netral. Setelah pelabelan selesai, model membuat kolom baru dalam kumpulan data berdasarkan model.