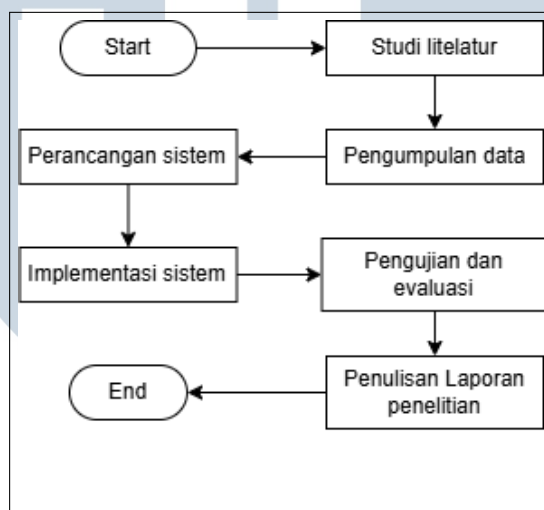


## BAB 3 METODOLOGI PENELITIAN

### 3.1 Gambaran Alur Penelitian

Dalam penelitian ini, dilakukan implementasi algoritma BERT untuk analisis sentimen pengguna Twitter terhadap ChatGPT. Berikut adalah penjabaran dari tahap penelitian sebagai berikut.



Gambar 3.1. Metodologi penelitian

### 3.2 Studi Literatur

Tahap ini dilakukan untuk memahami lebih lanjut terhadap penelitian yang akan dilakukan dengan mencari artikel dan jurnal yang berkaitan dengan topik penelitian. Jurnal dan artikel yang di ambil merupakan jurnal dan artikel nasional dan internasional yang diterbitkan dari rentang tahun 2019 sampai dengan 2023 dengan mencakup pembahasan permasalahan pada pengguna Twitter mengenai ChatGPT, Teori mengenai Analisis sentimen, Text Preprocessing, BERT, dan Confusion Matrix.

### 3.3 Pengumpulan Data

Untuk data Tweet tentang ChatGPT dari pengguna Twitter dikumpulkan dengan metode scraping dengan *keyword* "ChatGPT since:2024-12-01 until:2024-12-31 lang:id". Data yang didapat adalah tweet berbahasa Indonesia dari rentang

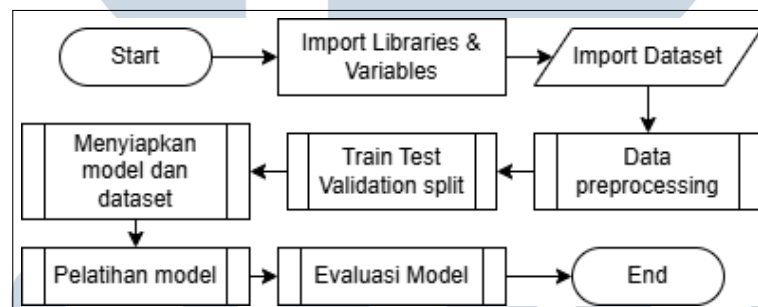
tanggal 1 Desember 2024 sampai dengan 31 Desember 2024 dengan jumlah 4.174 tweet, tweet yang sudah didapat kemudian disimpan ke dalam file berformat .CSV untuk diproses ke tahap selanjutnya.

### 3.4 Perancangan Sistem

Tahap ini dilakukan sebuah proses bagaimana tahap yang dilakukan untuk penelitian dalam mengklasifikasi Sentimen pada tweet mengenai ChatGPT. Untuk melakukan sebuah klasifikasi dilakukan beberapa tahapan yang perlu dilakukan.

#### 3.4.1 Flowchart Utama

Gambar 3.2 merupakan alur utama pada analisis sentimen yang dibuat. Dimulai dari import *library* dan variabel, memuat dataset, data preprocessing, data split, lalu menyiapkan model BERT untuk Training data dan evaluasi model.

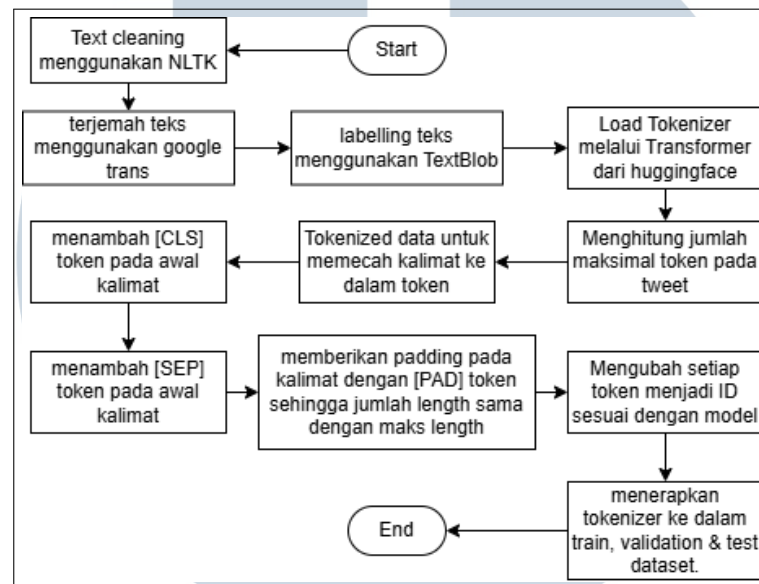


Gambar 3.2. Alur gambaran umum penelitian

#### 3.4.2 Flowchart Data Preprocessing

Gambar 3.3 merupakan alur pada data preproses. Tahap pertama yang dilakukan adalah *text cleaning* menggunakan NLTK, proses *text cleaning* mencakup pada *case folding*, *normalization* dan *tokenization*. Setelah *text cleaning* dilakukan, tahap selanjutnya adalah menerjemah isi dataset ke dalam bahasa inggris terlebih dahulu, karena tool untuk teks labelling hanya dapat membaca teks berbahasa inggris, setelah di terjemah, dilakukan proses teks labelling menggunakan TextBlob. Setelah melakukan labelling, dilakukan tokenisasi menggunakan BERT tokenizer melalui transformer yang didapat dari huggingface. Langkah yang dilakukan adalah load terlebih dahulu tokenizer, lalu menghitung jumlah maksimal token pada tweet, lalu tokenisasi data untuk memecah kalimat ke dalam token, lalu menambah CLS

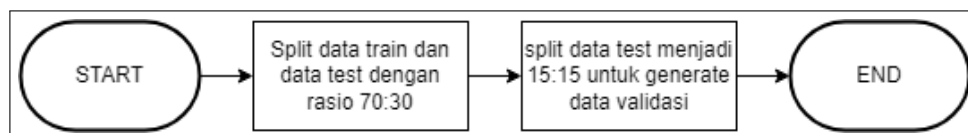
token pada awal kalimat, menambah SEP token pada awal kalimat, memberikan padding pada kalimat dengan PAD token sehingga jumlah length sama dengan maks length, Mengubah setiap token menjadi ID sesuai dengan model, dan kemudian menerapkan tokenizer ke dalam train, validation & test dataset.



Gambar 3.3. Flowchart tokenisasi data

### 3.4.3 Flowchart Pembagian Data

Pada tahap ini, data dibagi menjadi data train, test, dan validation. Proses ini dilakukan dengan menggunakan library scikitlearn. dikarenakan data terbagi menjadi 3, yang dilakukan adalah melakukan split data menjadi data train dan data sementara terlebih dahulu. Setelah data berhasil di split, data sementara di split kembali menjadi data validasi dan data test.

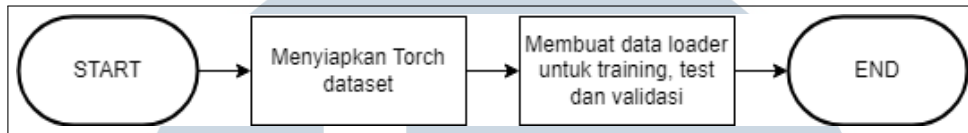


Gambar 3.4. Flowchart pembagian data

### 3.4.4 Flowchart persiapan model dan dataset

Gambar 3.5 merupakan tahapan untuk menyiapkan model dan dataset, Dataset harus di proses terlebih dahulu ke dalam PyTorch, sehingga teks dapat

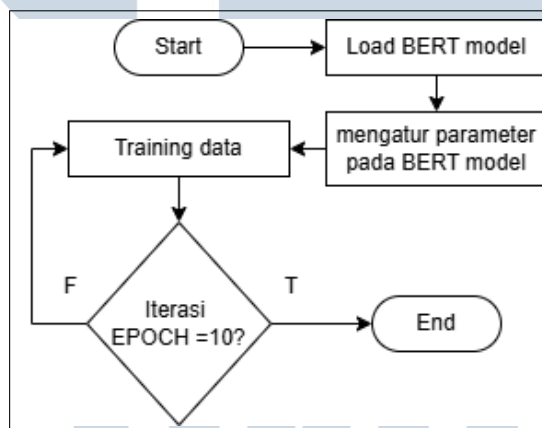
di return dan juga dapat di validasi hasil prediksinya, tahapannya terdiri dari menyiapkan torch dataset untuk memasukkan dataset ke dalam Pytorch, dan membuat data loader untuk *training, test, dan validation*.



Gambar 3.5. Flowchart Persiapan data model

### 3.4.5 Flowchart pelatihan data.

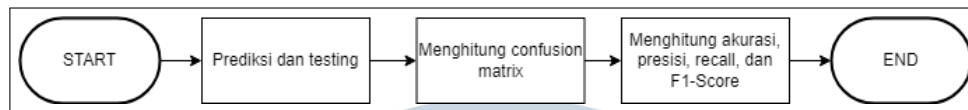
Pada tahap ini, Data yang sudah di tokenisasi akan dilakukan fine tuning, dengan cara load model BERT terlebih dahulu, lalu mengatur parameter pada model BERT, kemudian Fine tuning dijalankan. Gambar 3.6 merupakan alur pada persiapan model BERT dan fine tuning data.



Gambar 3.6. Flowchart pelatihan data

### 3.4.6 Flowchart Evaluasi data

Pada tahap ini, dilakukan evaluasi kinerja model BERT yang telah disesuaikan pada test set. Lalu passing data train yang telah diberi token melalui model dan mendapatkan label sentimen yang telah di prediksi. Lalu bandingkan label yang telah di prediksi dengan label dan lakukan perhitungan confusion matrix, dan juga metrik evaluasi akurasi, presisi, recall dan f1-score. Gambar 3.7 merupakan alur pada evaluasi data.



Gambar 3.7. Flowchart evaluasi data

### 3.5 Implementasi Sistem

Setelah melakukan perancangan sistem, penelitian ini masuk ke dalam tahap selanjutnya, yaitu membuat program untuk melakukan analisis sentimen dengan algoritma BERT, sesuai dengan flowchart yang sudah di buat di tahap sebelumnya.

### 3.6 Pengujian dan Evaluasi

Setelah membuat program untuk analisa sentimen menggunakan algoritma BERT, program akan di lakukan debugging untuk menemukan masalah yang terjadi pada program, seperti *bug* atau *error* pada program yang telah selesai dibuat. Setelah program dapat berjalan dengan sepenuhnya, dilakukan beberapa skenario pengujian dan evaluasi untuk menentukan hasil yang terbaik.

### 3.7 Penulisan laporan penelitian

Pada tahap ini, hasil penelitian didokumentasikan ke dalam bentuk laporan. Penulisan laporan penelitian mencakup teori yang didukung, perancangan, implementasi sistem hingga hasil pada penelitian. Tahap ini dilakukan agar penelitian dapat terdokumentasi dengan baik dan dapat digunakan sebagai bahan referensi untuk penelitian atau keperluan mendatang.

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